DOCUMENT RESUME

ED 055 447 EM 009 295

AUTHOR	Di Vesta, Francis J.; And Others
TITLE	Project Ikon: Studies of Imagery. Section I, Instructional Strategies: Multivariable Studies of Psychological Processes Related to Instruction.
INSTITUTION	Pennsylvania State Univ., University Park. Dept. of Educational Psychology.
SPONS AGENCY	Department of Defense, Washington, D.C. Advanced Research Projects Agency.
REPORT NO	ARPA-ORDER-1269
PUB DATE	Jul 71
NOTE	166p.; Annual Report, Part II. See also EM 009 296; EM 009 297
EDRS PRICE	MF-$0.65 HC-$6.58
DESCRIPTORS	Abstraction Levels; Imagery; Individual Differences; *Intermode Differences; *Learning Characteristics; Nonverbal Ability; *Pictorial Stimuli; Thought Processes; *Verbal Stimuli; Visual Learning; *Word Recognition
IDENTIFIERS	*Imagery Ability

ABSTRACT

Seven studies were conducted to investigate the interactions between imagery-ability and experimental treatments that parallel instructional procedures. The general orientation of the studies was to extend the concepts of imaginal and verbal coding systems to include individual differences in symbolic habits. Thus all the studies took into account the learner's ability to use the imaginal strategy, which was called imagery-ability, as the primary individual difference variable. The results of the studies suggest that imaginal processing is more effective for the learning of concrete words and verbal processing is more effective for the learning of abstract words. Another conclusion from these investigations is that imagery-ability reflects the ability of the learner to process the information by verbal or by imaginal strategies. The implications of these studies for the everyday process of education are that imagery-ability can be measured from tests involving manipulation of objects in space, that the acquisition of new concepts can be facilitated by providing direct experience prior to the provision of a verbal label, and that students with low imagery-abilities are severly handicapped when they are forced to employ imaginal rather than verbal strategies. (JY)

Project Ikon:

Instructional Strategies:

Multivariable Studies of Psychological Processes Related to Instruction

Francis J. Di Vesta | Nicholas M. Sanders

Charles B. Schultz | Paul D. Weener

Annual Report

July, 1971

Part II

Advanced Research Projects Agency

ARPA Order No. 1269

Monitored by

Office of Naval Research

Contract Number N00014-67-A-0385-0006

Department of Educational Psychology

College of Education

The Pennsylvania State University

ANNUAL REPORT

Part II

January 1, 1970 - June 30, 1971

ARPA Order Number:	1269
Program Code Number:	60949
Name of Contractor:	Pennsylvania State University University Park, Pennsylvania 16802
Effective Date of Contract:	1968 September 1
Contract Expiration Date:	1971 June 30
Contract Number:	N00014-67-A-0385-0006
Principal Investigator and Phone Number:	Francis J. Di Vesta 865-8303
Title of Work:	Instructional Strategies

Sponsored by
Advanced Research Projects Agency

ARPA Order No. 1269

TABLE OF CONTENTS

Section I

Project Ikon:

Studies of Imagery

Summary

Project Ikon: Studies of Imagery

Francis J. Di Vesta

In collaboration with

G. Susan Gray, Gary Ingersoll, Steven Ross, and Phyllis Sunshine

Technical Problem

This was a program of research consisting of seven studies in which were investigated the interactions between imagery-ability and experimental treatments that parallel instructional procedures. The general orientation was to extend the studies of imaginal and verbal coding systems of learning, operationally defined in terms of stimulus attributes (i.e., concreteness and abstractness) or instructional sets (i.e., imaginal vs. verbal processing) to include individual differences in symbolic habits. Thus, all of the present studies took into account the learner's ability to use the imaginal strategy, which was called imagery-ability, as the primary individual difference variable. Underlying this series of systematic studies was the reasoning that imagery-ability (i.e., high- and low-imagery) could be related to the attributes of stimuli (i.e., rated-imagery), or to the processes employed by the learner in acquiring, storing, and retrieving information (i.e., imaginal vs. verbal processing).

General Methodology

Initially, a series of three experiments were conducted in an attempt to replicate those conducted by Stewart (1965). In two of the studies the task materials were presented pictorially and verbally in contrastive treatments. In one of these two studies we examined the ability of high- and low-imagers to transfer material, learned via pictures and verbalizations, to new situations while in the other we examined the learner's ability to recall material presented via the two methods. In the third replication study, the influence of concrete and abstract verbal stimuli (rather than pictures and words) as variables affecting rate of acquisition and recall was investigated. A factor analytic study attempted to clarify the relationship between subjective reports and objective tests of imagery ability. In a fifth study we investigated, experimentally, the interaction between pictorial and verbal contexts of material to be learned on subsequent ability to transfer to new situations by high- and low-imagers. A sixth study was designed to examine the effect of imagery-ability in acquiring information from tasks in which the noun-adjective relationship in paired-associates was varied according to relationship between stimuli and responses (i.e., noun-adjective vs. adjective-noun sequence) and concreteness of the stimulus or response. In the seventh study the effectiveness of verbal and imaginal processing by high- and low-imagers was investigated.

Technical Results

In general, the results provided support for Paivio's (1970) two-stage model of associative learning as follows: The meanings of concrete words are learned via direct experience and intraverbal experience. Accordingly, they evoke both images and verbal responses. The meanings of abstract words, on the other hand, are learned primarily by association with other words, and, hence, elicit primarily verbal responses. Thus, imaginal processing is more effective for the processing of concrete words and verbal processing is more effective for the processing of abstract words.

Another conclusion, from these investigations, was that imagery-ability, as measured in these studies, reflects the ability of the learner to process the information by verbal or by imaginal strategies. There was little or no evidence that imagery ability reflected a sensitivity on the part of the learner to profit from pictorial or verbal stimuli, per se, as suggested by Stewart's (1965) studies. Nevertheless, the acquisition of all learners was facilitated more by pictorial than by verbal presentations. An interesting adjunct to this conclusion was that the picture-word order was always more favorable for learning, transfer, and recall than was the word-picture order. Furthermore, over several trials, varying the modality (e.g., presentation of a picture, then its verbal label, and then the picture again) facilitated recall to a greater extent than retaining the same modality (e.g., presentation of a picture) on all trials. Finally, the factor analytic study indicated that the verbal-ability and imagery-ability factors were orthogonal rather than bi-polar. Imagery tests based on subjective reports of ability to use imaginal processing was related primarily to social desirability rather than to objective tests of imagery ability. This latter finding undoubtedly accounts for some of the failures to identify relationships between individual differences in imagery-ability and ability of the person to profit from pictorial versus verbal materials either in terms of rate of acquisition or in terms of the facilitation of memory, but is not an exclusive reason.

Educational Implications

The implications of the results of these studies are, in all cases, more or less self-evident, if we are permitted to extrapolate from experimental to classroom settings. Thus, to mention a few implications: Imagery-ability can be measured from tests involving manipulation of objects in space. The acquisition of new concepts can be facilitated by providing direct experience prior to the provision of a label. Variations in the modality of stimuli tends to facilitate the recall of the material learned. Finally, imagery-ability means just that; high-imagers do use imaginal processing more effectively than do low-imagers. Conversely, low-imagers are severely handicapped when they are forced to employ imaginal strategies rather than verbal strategies. This finding implies that in adapting to this individual difference, high-imagers can be taught by methods which capitalize on imaginal processing (e.g., instruction which involves graphic displays) while low-imagers might be taught by methods which employ verbal processing (e.g., instruction via lecture methods).

Implications for Further Research

These studies indicate that the study of aptitude by treatment interactions with imagery would be most fruitful if the research strategy were to emphasize the nature of imaginal processing. Certainly, more needs to be known about the effects of the pictorial-labeling sequence in instructional strategy. However, as important is an understanding of the differences between verbal and imaginal processing. It would appear that the techniques employed in the present studies, with but some slight modification such as time-sampling could provide further insights into the way materials are transformed by the subject.

References

Paivio, A. On the functional significance of imagery. Psychological Bulletin, 1970, 6, 385-392.

Stewart, J. C. An experimental investigation of imagery. Unpublished doctoral dissertation, University of Toronto, 1965.

The Recognition and Recall by High and Low Imagers of Stimuli Presented as Words and as Pictures

Francis J. Di Vesta

In an investigation by Jenkins (1963) as cited and described by Stewart (1965) Ss were presented a series of pictures and words. Then, on a subsequent task, the Ss were presented either pictures, words, words associated with pictures seen on the first presentation, and pictures associated with words seen on the first presentation. The S's task was to indicate which of the items in the second series he had seen on the first series. The results demonstrated that pictures were easier to recognize than words, or stated conversely, more errors were made in recognizing words than were made in recognizing pictures. In addition, pictures were mistaken for words less often than words were mistaken for pictures seen before. These findings suggest the greater generalization of words over pictures. Indeed, the authors labeled the tendency to make more errors with words than with pictures as a case of response generalization.

In a follow-up study, Stewart (1965) modified the procedures used by Jenkins, et al. and extended their study by investigating the differences in performance on the recognition task between high and low imagers. The individual differences were defined by scores on the Spatial Relations sub-test of the Differential Aptitude Test Battery (Bennett, Seashore, & Wesman, 1963) and by the Space Thinking (Flags) Test (Thurstone and

Jeffrey, 1959). Stewart summarized her results as follows:

" . . . The pictures were recognized with significantly fewer errors than words. Presenting the items as pictures benefited the high imagers to a greater extent than it did the low imagers; at the same time, both groups were aided significantly [by the pictorial representations]. There was <u>some</u> [emphasis ours] evidence that high imagers were more likely to code a word as a picture than were the low imagers; and vice versa, the low imagers were more likely than were the high imagers to remember or code a picture as a word" (Stewart, 1965, p. 74).

The importance of these findings is in the suggestion that learning materials tend to be coded in the same form they were received by the S. In addition, recognition based on percepts appears to be easier than that based on symbols. Furthermore, there appears to be a tendency for high and low imagers to code materials in different ways, thereby also affecting the retrieval of information. Nevertheless, the theoretical implications of Stewart's results for retrieval are far from clear.

Because of its implications for understanding the processing of incoming information as well as for understanding the interaction between aptitude and individual differences, Stewart's study was replicated as well as extended in the present study. The first phase of the S's participation was identical to that required of Ss in Stewart's study. That is, following an initial presentation of the word and picture stimuli, the materials were presented again, in varying relationships to those presented originally, and the S's task was to identify those he could recognize as having been presented before. In the second task, which immediately followed the recognition series, the materials on the first task were again presented after which the S was required to recall as many of the items presented in any order he chose; that is the free recall procedure was used. As a replication of Stewart's study the recognition phase of the present study was intended to provide another test of the hypothesis that the coding basis is the form in which the stimuli are presented (i.e., most of the correct responses were expected to be in the form they were presented initially). As an extension of Stewart's study, the present study permitted an examination of the hypothesis that material is retrieved according to the dominant basis of encoding as it is influenced by whichever strategy is reflected in his imagery scores. Thus, it was expected that pictorial material would be encoded more easily than verbal material by high imagers. While the same effects were hypothesized for low imagers (because evidence from earlier studies suggest that pictorial stimuli are learned and recalled more easily than verbal stimuli) it was expected that the difference would not be as great as it would be for high imagers. Furthermore, it was expected that the retrieval preferences of high imagers would be reflected in clustering during recall of pictorial materials to a greater extent than of verbal materials.

Method

Design

The Ss were administered a list of 50 words randomly assorted with 50 pictures representing common objects (Presentation Trial: I). They were then presented another list (Recognition Trial) consisting of 25 words seen before (WW), 25 pictures seen before (PP), 25 words seen as pictures (PW), 25 pictures seen as words (WP), 25 pictures never seen, either as pictures or words, on the presentation trial (NP), and 25 words never seen, either as pictures or words, on the presentation trial (NW). The Ss task was to indicate which items they had seen before, as pictures or as words, and which items had not appeared before. They were then readministered the first list (Presentation Trial: II) after which they were given a 5-min. recall period in which they were to record as many items as possible via free-recall, from the Presentation list. The primary, though not exclusive, analyses were made of PP and WW items correctly recalled in the Recognition Trial and of PPP and WWW items correctly recalled in the Recall Trial. Sex of subject and levels of imagery were included as variables in some of the analyses.

Subjects

One hundred and four female and 80 male educational psychology undergraduates served as Ss in the experiment. From this group the data for groups (n = 25) of men and women high- and low-imagers were selected for analysis. All Ss were informed that the present study was one of several related experiments and that attendance at all sessions was mandatory if credit toward the course grade was to be received.

Measures of Individual Differences

A battery of tests designed to measure imagery ability, verbal ability, and automatization was administered to the entire group of Ss. The tests for measuring imagery consisted of the Flags Test (Thurstone & Jeffrey, 1959), the Spatial Relations Test (Bennett, et al., 1963), and the Gottschaldt Figures Test as described by Thurstone (1944). Verbal ability was measured by a vocabulary test (designed especially for this study by selecting items from existing intelligence tests) and by the College Entrance Examination Board Scholastic Aptitude Test: Verbal

(1962-1963). Automatization was measured by the Stroop Color-Word Test as described by Thurstone (1944) and by an automatization test described by Broverman, Klaiber, Kobayashi, and Vogel (1968). The socres on these tests were standardized for the groups of males and females separately and the resulting T scores for tests comprising a given factor were summed and averaged. The data for the 25 males and 25 females with high scores and the 25 males and 25 females with low socres on these factors were employed in analyses involving aptitude by treatment interactions. The average raw scores for these groups are presented in Table 1. More detailed information on these tests and relationships with other tests administered in the same battery are described in a report by Di Vesta, Ingersoll, and Sunshine (1971).

The analyses presented in this paper are based only on high and low imagers. Other analyses based on high and low automatizers and high and low verbal ability had also been made, In addition, analyses had been made of high and low imagers selected according to the procedure described by Stewart (1965) using only the Flags and Spatial Relations Tests. However, these latter analyses either yielded no effects of consequence or added no further information to the results of the analyses based on the factor scores for imagery. Accordingly, the results of those analyses are not reported here.

Materials

A pool of 300 words from Stewart's (1965) study was used as a basis for preparing stimulus materials. All words were common concrete nouns. Each word had a picture as its counterpart. There were three 100 item lists prepared for Presentation Trial: I. Each list consisted of 50

Table 1

Means of Raw Scores For Men and Women on Tests Comprising Imagery, Verbal, and Automatization Factors in High and Low Groups for Each Factor[a]

Tests	Women		Men	
	Low	High	Low	High
Imagery Factor[b]				
Flags	71.17	121.35	92.64	123.40
Spatial Relations	49.33	88.30	54.00	92.52
Gottschaldt (Total)	27.87	44.35	27.16	46.28
Vocabulary	20.08	18.52	15.88	19.44
SAT (Verbal)	552.23	517.75	478.81	509.45
Verbal Factor				
Vocabulary	15.28	22.16	14.12	21.48
Reading	13.04	19.60	14.24	19.68
SAT (Verbal)	441.64	581.60	441.60	557.24
Automatization[b]				
Stroop (Color-word score)	77.71	114.00	85.76	129.92
Automatization	46.33	62.76	48.64	62.48
Vocabulary	20.86	18.00	18.76	17.16
SAT (Verbal)	572.83	492.64	515.50	500.61

[a] N = 25 in each group.

[b] Vocabulary and SAT (Verbal) scores are presented for these groups to indicate any relationships with the factor tests. Only the differences between high and low automatizers (women) on SAT Verbal scores were significant ($p < .01$).

picture and 50 word stimuli compiled by the random selection of items without replacement from the total pool of words or pictures. Within each list for the Presentation Trial no picture was represented by a word duplicating it. This procedure resulted in three unique lists.

Each of the three lists for the Recognition Trial was constructed to correspond with one of the three series for the Presentation Trial: I. A given list consisted of 150 stimuli, 75 of which were pictures and 75 were words. The items were selected so that half (n = 25) of the word stimuli used in the Presentation Trial: I were maintained as words for the Recognition Trial while the other half (n = 25) were represented pictorially. Similarly, half (n = 25) of the original pictures were maintained as pictures for the Recognition Trial while the other half (n = 25) were represented as words. The remaining 50 items were divided equally between words and pictures none of which had either verbal or pictorial counterparts in the rest of the list; nor had these 50 items been seen, in either form, on Presentation Trial: I. Thus, each Recognition list consisted of 50 stimuli (25 words, 25 pictures) in the form originally seen on the Presentation Trial: I, 50 stimuli in converted (associated) form, and 50 stimuli previously unseen (new), 25 of which were pictures and 25 of which were words. The stimuli were ordered randomly, via reference to a table of random digits, within the lists.

Procedure

The task was administered to Ss in groups varying in number from 2 - 8. However, each S worked individually and at separate locations (every other seat) in a room approximately 3 x 5 meters in size. The Ss were seated opposite a large screen and were informed that they would be participating in an experiment on memory. They were told that a series of slides, showing pictures and words, would be projected onto the screen and that their task would be to study the slides and attempt to remember as many items as possible. Following these instructions, one of the three Presentation Trial lists, selected at random, was projected at a rate of 3-sec. per slide.

Immediately after Presentation Trial: I the Ss were provided with an answer sheet on which were spaces for indicating whether a slide to be presented on the Recognition Trial had been seen on Presentation Trial: I and, if so, whether it had been seen as a picture or a word. Each slide on the Recognition trial was presented at a rate of 8-sec. per slide. The S was informed that his task on the Recognition Trial was to decide whether the item on the slide a) was completely new, i.e., it was not seen in the presentation series either as a picture or a word; b) was seen previously as a word or; c) was seen previously as a picture, and to mark his (or her) answers accordingly on the answer sheet. The Ss were not told the proportion of slides represented in each category. Although guessing was encouraged where the S was uncertain, E also expressed the need for rapid and accurate decisions. In order to aid the S in keeping his place, the slides for the Recognition Trial were ordered numerically and E called out the number of each slide as it appeared on the screen.

After the Recognition Trial was completed, Presentation Trial: II, consisting of the same series as Presentation Trial: I, was administered. The Ss were instructed to study each item carefully and to remember as many items as possible. At the completion of the slide presentation, blank sheets of paper were distributed for the Free-Recall of all items (whether picture or word) that the S could remember from Presentation Trial: II and in any order he chose. The Ss were not required in the Recall Phase to indicate whether the item recalled had been presented as a picture or a word during Presentation Trial: II.

Results

The primary dependent variables, based on responses to the Recognition Trial, were: a) the number of items recalled and labeled correctly; and b) the number of items said to have occurred as words or pictures on Presentation Trial: I when in fact they had not been presented at all. The dependent variable based on responses to the Free-Recall task was the number of items, presented as words or pictures on Presentations I and II, that were recalled correctly. Pairs of pictures or of words, scored as adjacencies, were used as measures of clustering in further analyses.

All analyses of data were made by a mixed analysis of variance design. However, the variables used differed from one analysis to another. Typically, the between-subjects variables were the randomizations of the items (Item Series), Imagery Ability, and Sex of subject. The within-subjects variables for analyses of responses made on the Recognition Trial consisted of the Form of the Stimuli on the Presentation Trial: I and the Form of Stimuli on the Recognition Trial. In addition, the analyses of responses on the Free-Recall Task included the form of the stimuli on the Presentation Trial: II as another within-subjects variable. Inasmuch as it was virtually impossible to interpret significant main effects or interactions involving the Item Series, their occurrence will be noted in the reporting of results without further comment.

Recognition

The analyses of the number of items correctly recognized on the Recognition Trial by female High and Low Imagers are summarized in Table 2. The parallel analysis for the male High and Low Imagers are summarized in Table 3. In both tables comparisons have been made between combinations of words and pictures shown on Presentation Trial: I and the Recognition Trial. Thus, words on both trials (WW) have been compared with pictures on both trials (PP); new words (NW) of objects (i.e., a new set of words was presented on the Recognition Trial the items of which had not appeared either as words or as pictures on the Presentation I trial) were compared with new pictures (NP) on the Recognition Trial; and words on Presentation Trial: I followed by pictures on the Recognition Trial (WP) were compared with pictures followed by words (PW).

In these analyses the main effects of Item Series was significant only for the comparison of responses by male Ss to the WW and PP stimuli (F [2,54] = 3.84, $p < .05$). However, Item Series interacted with Imagery Level in the comparison of responses by female Ss to the WW and PP stimuli (F [2,54] = 5.04, $p < .05$) with Form of Stimuli in the comparison of responses by male Ss to the NW and NP stimuli (F [2,54] = 4.90, $p < .05$).

The most important and clear cut effects to be noted in these results, however, are in the comparisons between Form of Stimuli. In all analyses (see Table 2) the differences were highly significant ($p < .001$). As shown in Table 4, for both men and women Ss, performance on PP stimulus combinations was more accurate than on the WW combinations; performance

20

Table 2

Summary of Analysis of Variance of the Number of Items Correctly Recognized On the Recognition Trial by Women High and Low Imagers

		Comparison					
		WW vs. PP		NW vs. NP		WP vs. PW	
Between Subjects	df	MS	F	MS	F	MS	F
Item Series (A)	2	20.03	1.00	.41	.02	10.68	0.55
Imagery Level (B)	1	3.68	0.18	2.13	.08	7.01	0.36
A x B	2	100.80	5.04[b]	4.35	.17	32.60	1.69
Error (b)	54	20.02		25.90		19.26	
Within Subjects							
Form of Stimuli (C)	1	516.68	82.36[d]	163.33	19.81[d]	156.41	14.45[d]
A x C	2	12.70	2.03	36.26	4.40[b]	9.16	0.85
B x C	1	6.08	0.97	10.80	1.31	9.08	0.84
A x B x C	2	7.30	1.16	1.08	0.13	9.03	0.83
Error (w)	54	6.27		8.24		10.83	

[a] $p < .10$

[b] $p < .05$

[c] $p < .01$

[d] $p < .001$

Table 3

Summary of Analysis of Variance of the Number of Items Correctly Recognized on the Recognition Trial by Male High and Low Imagers

		Comparison					
		WW vs. PP		NW vs. NP		WP vs. PW	
Between Subjects	df	MS	F	MS	F	MS	F
Item Series (A)	2	99.99	3.84[b]	7.06	0.19	49.03	1.56
Imagery Level (B)	1	2.41	0.09	33.21	0.97	116.03	3.69[a]
A x B	2	12.86	0.49	87.86	2.41	21.73	0.69
Error (b)	54	26.03		36.43		31.43	
Within Subjects							
Form of Stimuli	1	715.41	56.85[d]	69.01	10.09[d]	145.20	15.21[d]
A x C	2	26.66	2.12	35.51	4.90[b]	.40	0.04
B x C	1	1.86	0.15	.01	0.00	3.33	0.35
A x B x C	2	8.18	0.65	7.11	1.04	10.03	1.05
Error (w)	54	12.58		6.84		9.55	

[a] $p < .10$

[b] $p < .05$

[c] $p < .01$

[d] $p < .001$

22

Table 4

Mean Numbers of Correct Responses on Recognition Trial

Form of Stimuli on Presentation I and Recognition Trials	Sex of Subject		
	Men	Women	Row Means
Word-Word	15.83	17.08	16.46
Word-Picture	15.72	17.23	16.48
Picture-Word	17.92	19.52	18.72
Picture-Picture	20.72	21.23	20.48
None-Word	17.87	18.83	18.35
None-Picture	16.35	16.50	16.43

on the PW order of presentation was clearly better than on the WP order; and new words (NW) were recognized more accurately as not having appeared on the presentation trial than were new pictures (NP). In every case, without exception, the performance of women on these tasks exceeded the accuracy of the performance of men.

Similar analyses of recognition errors are summarized in Table 5. The reader should note that these analyses are not of mere incorrect recognitions which would be only the difference between actual and possible correct responses. Instead, this analysis is of the errors made in labeling a stimulus on the Recognition Task as having been presented as a word or picture or Presentation Trial: I or as not having been seen on Presentation Trial: I. If a picture on the Recognition Trial was said to have been presented as a picture on Presentation Trial: I, when indeed, it had been presented as a word, this was called a word-picture error (WPER). If a word presented on the Recognition Trial was said to have been presented as a word on Presentation Trial: I when it actually had been presented initially as a picture, this was called a picture-word error (PWER). NWER and NPER refer to errors in which a word or picture appearing on the Recognition Trial were incorrectly recognized as having been presented on Presentation Trial: I. Thus, NWER errors refer to the number of times an S said that a word had appeared on Presentation Trial: I as either a word or picture when it had not been so presented; conversely, NPER errors refer to the number of times an S said that a picture had appeared on the Presentation Trial: I as either a word or picture when it had not been so presented.

The analyses summarized in Table 4 yielded significant main effects due only to comparisons between Form of Stimuli. The comparison of WPER

Table 5

Summary of Analyses of Variance of Errors Made on Recognition Trial

Source of Variance		Word-Picture (WP) Errors vs. Picture Word (PW) Errors				New Words (NW) vs. New Picture (NP) Errors			
		Females		Males		Females		Males	
Between Subjects	df	MS	F	MS	F	MS	F	MS	F
Item Series (A)	2	.75	.05	38.43	1.36	.41	.02	7.06	.19
Imagery Level (B)	1	7.50	.50	13.33	.47	2.13	.08	35.21	.97
A x B	2	25.98	1.74	34.53	1.22	4.36	.17	87.85	2.41
Error (b)	54	14.91		28.21		25.91		36.43	
Within Subjects									
Form of Stimuli (C)	1	22.53	5.75[b]	3.33	.36	986.13	62.34[d]	1122.42	64.96[d]
A x C	2	8.66	2.21	38.03	4.10	22.16	1.40	59.16	3.42
B x C	1	1.63	.41	6.53	.70	8.53	.54	2.41	.14
A x B x C	2	.51	.13	4.93	.53	15.41	.97	4.65	.27
Error (w)	54	3.92		9.28		15.82		17.28	

with PWER errors yielded F (1,54) = 5.75, $p < .05$ for data based on responses of female Ss. Though this main effect was not significant for male Ss, there is some indication that the Item Series may have somehow been influential in this regard inasmuch as the interaction of Item Series and Form of Stimuli yielded F (2,54) = 4.10, $p < .05$ for data based on male Ss. The comparison of NWER and NPER were clearly the same for female and male Ss since the analyses yielded F (1,54) = 62.34 and F (1,54) = 64.96, respectively, both of which are highly significant ($p < .001$).

The means for these comparisons are presented in Table 6. There it may be seen that there were relatively more errors caused by mislabeling a word on the Recognition Trial when it had been presented initially as a picture, than in mislabeling a picture on the Recognition Trial when it had been presented as a word on Presentation Trial: I. A major error of the Ss was to indicate that a word presented on the Recognition Trial had been presented on the Presentation Trial: I when it had not been. The data related to these findings support those reported by Stewart.

Free Recall

The analyses of variance of the number of items correctly recalled on the Free-Recall task are summarized in Table 7 for the data based on the responses of female Ss and in Table 8 for data based on the responses of male Ss.

The result of primary interest in these analyses is the effect due to the interaction of Imagery Levels with Form of Stimulus on Presentation Trial: I. This effect yielded F (2,108) = 3.10, $p < .05$ for the data

Table 6

Mean Number of False Recognitions on the Recognition Trial

by Men and Women High and Low Imagers

Presented on Presentation Trial: I and Recognition Trial as:	Described on Recognition Trial as having been presented on Presentation I as:	Low Imagers		High Imagers		Total	
		Men	Women	Men	Women	Men	Women
Word-Picture	Picture	5.17	2.90	4.03	3.17	4.60	3.03
Picture-Word	Word	5.03	3.53	4.83	4.27	4.93	3.90
Not Presented	Word	11.63	9.80	10.27	10.60	10.95	10.20
Not Presented	Picture	5.23	4.60	4.43	4.33	4.83	4.47

27

Table 7

Summary of Analysis of Variance of the Number of Items Correctly Recalled on the Free-Recall Task by Female High and Low Imagers

	df	MS	F	
Between Subjects				
Item Series (A)	2	3.85	0.35	
Imagery Level (B)	1	25.60	2.30	
A x B	2	1.61	0.15	
Error (b)	54	11.11		
Within Subjects				
Form of Stimuli on Presentation I (C)	2	3561.20	602.74	$p < .001$
A x C	4	47.97	8.12	$p < .01$
B x C	2	18.31	3.10	$p < .05$
A x B x C	4	1.58	0.27	
Error (w_1)	108	5.91		
Form of Stimuli on Presentation II (D)	1	5.38	1.43	
A x D	2	18.04	4.78	$p < .05$
B x D	1	1.88	0.50	
A x B x D	2	1.64	0.43	
Error (w_2)	54	3.77		
C x D	2	18.43	2.97	$p < .10$
A x C x D	4	24.87	4.01	$p < .01$
B x C x D	2	4.80	0.78	
A x B x C x D	4	1.90	0.31	
Error (w_3)	108	6.20		

Table 8

Summary of Analysis of Variance of the Number of Items Correctly Recalled on the Free-Recall Task by Male High and Low Imagers

Between Subjects	df	MS	F	
Item Series (A)	2	14.17	1.27	
Imagery Level (B)	1	96.10	8.61	$p < .01$
A x B	2	.98	0.09	
Error (b)	54	11.16		
Within Subjects				
Form of Stimuli on Presentation I (C)	2	2786.32	391.47	$p < .001$
A x C	4	31.77	4.46	$p < .01$
B x C	2	47.11	6.66	$p < .01$
A x B x C	4	8.77	1.16	
Error (w_1)	108	7.12		
Form of Stimuli on Presentation II (D)	1	1.11	0.28	
A x D	2	8.30	2.08	
B x D	1	2.18	0.55	
A x B x D	2	7.15	1.79	
Error (w_2)	54	3.99		
C x D	2	8.05	1.75	
A x C x D	4	3.82	8.31	
B x C x D	2	.29	0.06	
A x B x C x D	4	.92	0.20	
Error (w_3)	108	4.60		

based on the responses of women Ss and F (2,108) = 6.66, $p < .01$ for the data based on the responses of men. In addition, the main effect associated with Imagery Levels for men yielded F (1,54) = 8.61, $p < .01$.

In Table 9 it can be seen that the women recalled more items than men after Presentation Trial: II. However, of more interest is the finding that recall a) is directly related to the form of stimuli on the two presentations and b) is affected by the form of the stimuli on the recognition trial. Thus, W-W presentation (i.e., word on Presentation Trial: I and Presentation Trial: II, disregarding the form on the Recognition Trial) results in poorer performance than a P-P presentation as can be seen in Table 10. However, either W-W or P-P presentations, when interpolated with the other form on the Recognition Trial (thus, either WPW or PWP) results in better free-recall than when the same form is used in the Recognition Trial (that is, either WWW or PPP).

The data presented in Table 11 are related to the interaction of Imagery Levels with the Form of Stimuli on the Presentation Trials. There it can be seen that High Imagers (both men and women) recall more items, in general, than do Low Imagers. However, in both groups (that is, men and women) High Imagers recall more items depicted pictorially than do the Low Imagers. These data, however, must be interpreted cautiously. The direct relationship between the difference in word and picture recall and the total recall scores (or either word or picture recall scores singly) suggests that Imagery Levels may be reflecting a general intellectual factor. Consequently, the data may simply imply that the person with the higher ability is able to profit more from the optimal treatment which, presumably, is the pictorial presentation of stimuli.

Table 9

Mean Numbers of Correct Responses Recalled on Free-Recall Trial

As a Function of the Effect Due to Form of Stimuli on Presentation Trials

Form of Stimuli on Presentation Trials	Sex of Subject	
	Men	Women
Words	8.78	9.71
Pictures	10.08	11.24
New*	1.15	1.13

* These are intrusions from the Recognition Trial. These words <u>or</u> pictures were present on the Recognition Trial only and were not present on either Presentations I and II.

Table 10

Mean Number of Correct Responses on Free Recall Trial

Form of Stimuli on Presentation Trial: I, Recognition Trial, and Presentation Trial: II	Sex of Subject	
	Men	Women
Word-Word-Word	8.55	9.20
Word-Picture-Word	9.02	10.21
Picture-Picture-Picture	9.83	10.97
Picture-Word-Picture	10.32	11.51
None-Word-None	.98	1.00
None-Picture-None	1.33	1.27

32

Table 11

Mean Number of Items Correctly Recalled by Men and Women High and Low Imagers On Free-Recall as a Function of Form of Stimuli on Presentation Trials

Form of Stimuli	High Imagers			Low Imagers		
on Presentation Trials	Men	Women	Total	Men	Women	Total
Words	9.63	10.08	9.86	7.93	9.33	8.63
Pictures	10.98	11.83	11.40	9.17	10.65	9.91
Not Presented	0.95	0.97	0.96	1.37	1.30	1.33
Diff (p-w)	1.35	1.75	1.55	1.24	1.32	1.28

A final analysis was made of the proportion of adjacent pictures or words in the free-recall task. These were computed simply by dividing the number of observed pairs of pictures or words by the number of possible pairs of pictures or words respectively. None of the main effects or interactions involving these data were significant ($p > .05$). However, since the trends for the interaction between Imagery Level and Form of Adjacent Pairs were very similar to those reported for similar data by Stewart (1965) they are presented in Table 12. The reader should note that this interaction is based on data for all Ss and not just extreme groups used in previous analysis. The interaction yielded F (1,116) = 2.63, $p < .20 > .10$. If this finding can be considered a reliable one since it replicates Stewart's findings, it implies that Low Imagers tend to organize (retrieve) materials presented in verbal form more efficiently than materials presented in pictorial form. On the other hand, the data for high Imagers imply greater organization of pictorially presented materials than of verbally presented material.

Discussion

This experiment has clearly replicated the earlier findings by Stewart (1965) with regard to the recall of materials presented in picture and word forms. In addition, it provides clear evidence for the differential performance of men and women on the kinds of tasks that were used. On the other hand, though there was some slight evidence that Ss classified as high and low Imagers perform differently on materials presented in picture or word forms the data can only be considered as suggestive.

Table 12

Proportions (Observed/Possible) of Adjacent Pictures or Words

in the Free-Recall of High and Low Imagers

	Proportion of	
Imagery Level	Word Pairs	Picture Pairs
Low	.14	.12
High	.12	.18

35

Perhaps the most important implication of the present study is to be obtained from an integration of the findings from the recognition and free-recall phases which suggest the relative roles of imaging and labeling in storage and retrieval processes in memory. Since pictures were more frequently recalled than words there is the implication that words, as highly generalized symbols, refer to idealized attributes or qualities and consequently create more interference in recall than do pictures. Speculation regarding the specificity of the referent for pictures has been made in the introduction to this paper. Since pictures are relatively more distinctive than words it can be assumed they will suffer less from interference during the Recognition Trial.

The findings also clearly imply that the picture-to-word order results in more correct identifications than the word-to-picture order. This effect may be the result of factors associated with developmental processes, i.e., Ss in the college culture have had more experience in providing verbal labels for pictures than in providing images for words. Nevertheless, for whatever reason, the finding that providing a label for a picture does facilitate recognition more than providing a picture for the word is a clear and reliable one. It would appear, by way of explanation, that the verbal response was included in responses to the picture but the percept or imaginal (picture) response was less likely to be included in responses to the verbal label for the referent. This explanation indicates that redundancy on the Presentation, Recognition, and Recall Trials (e.g., word-word or picture-picture) is not as efficient for recall as is bi-modal presentation. These explanations also apply to the findings that new words are more easily recognized than new pictures. Thus, because there are fewer words and more pictures incorporated into the S's memory during the first presentation, there will be less interference to new words than to new pictures.

In general, the findings from the present study imply that materials are received first as images and then verbally labeled for storage. If this assumption is correct, it would also be expected that the picture-word order of presentation would yield more correct responses in free recall than the word-picture order of presentation, because the latter requires S to perform an additional operation of reversing the order. The picture-word order should also be superior to the picture-picture order which does not provide for labeling thereby hindering effective storage, or the word-word order which does not provide a percept to make the label easily discriminable (less generalized). These assumptions were provided substantial support in this study. The findings are especially interesting since the PPP was not the most efficient order of presentation as would have been suggested if only the Ss' performance on the recognition trial had been investigated.

The aforementioned findings and assumptions imply a dual process in retrieval of information. On the one hand, recognition depends on the distinctiveness of percepts, in which case the pictures are more easily identified than words. On the other hand, retrieval depends on encoding processes or the strategy by which materials are stored, in which case the picture which has been labeled by a word is more easily retrieved than words followed by words (which are subject to interference); or than words followed by pictures because pictures presumably add little to the distinctiveness of a generalized symbol; or than pictures followed by pictures because images or percepts are retained only for brief periods of time unless they are labeled.

The superiority of the female Ss recall over the males recall is undoubtedly due to factors associated with developmental processes. It is too early to say what the nature of the factors that account for this differential performance might be.

None of the results of the present study provided support for Stewart's (1965) findings which she summarizes as follows:

"High imagers were superior to low imagers in picture recognition but inferior in word recognition. Though the picture recognition was significantly better for both types of imagers than was the word recognition, the high imagers benefitted to a greater extent."

Part of the reason for lack of supporting evidence may have been that the Ss in the present study achieved higher mean recognition and recall scores than those in Stewart's study. Although the interactions associated with Imagery Ability on the Free-Recall task implied that high Imagers benefited more by pictures than by words, the alternative explanation remained that such differences might be attributable to other ability factors associated with Imagery. Furthermore, while the finding that low Imagers tend to organize by words in free-recall and high Imagers tend to organize by pictures supports the tendencies of the two groups suggested in Stewart's study, the differences were not significant.

Thus, we are led to conclude that the present results point to a clear superiority of pictures over words for presentation of learning material related to a given referent if that material is to be recognized easily on later occasions. Furthermore, the superiority of the picture to word over the word to picture order suggests that percepts are provided verbal labels for storage and that such labels facilitate later retrieval as measured by the number of items recalled on a free recall task. If these treatment variables interact with Imagery levels to affect recognition or recall, the effect must be considered to be a fragile one or of limited generalizability at best.

References

Bennett, G. K., Seashore, H. G., & Wesman, A. G. Differential aptitude tests. (Grades 8-13 and adults.) New York: Psychological Corporation, 1963.

Broverman, D. M., Klaiber, E. L., Kobayashi, Y., & Vogel, W. Roles of activation and inhibition in sex differences in cognitive abilities. Psychological Review, 1968, 75, 23-50.

College Entrance Examination Board Scholastic Aptitude Test. Princeton, N. J.: Educational Testing Service, 1962-1963.

Di Vesta, F. J., Ingersoll, G., & Sunshine, P. A factor analysis of imagery tests. Journal of Verbal Learning and Verbal Behavior, 1971, in press.

Jenkins, J., Carlson, Tweedy and Doren. Response generalization. University of Minnesota, Unpublished manuscript, presented at the Second California Conference on Verbal Learning and Verbal Behavior, 1963.

Lockhart, R. S. Retrieval asymmetry in the recall of adjectives and nouns. Journal of Experimental Psychology, 1969, 79, 12-17.

Paivio, A. Learning of adjective-noun paired-associates as a function of adjective-noun word order and noun abstractness. Canadian Journal of Psychology, 1963, 17, 370-379.

Stewart, J. C. An experimental investigation of imagery. Unpublished doctoral dissertation, The University of Toronto, 1965.

Thurstone, L. L. A factorial study of perception. Chicago: University of Chicago Press, 1944.

Thurstone, L. L., & Jeffrey, T. G. Space thinking (Flags). Chicago; Education-Industry Service, 1959.

Yuille, J. C., Paivio, A., & Lambert, W. E. Noun and adjective imagery and order in paired-associate learning by French and English subjects. Canadian Journal of Psychology, 1969, 23, 259-466.

Verbal and Imaginal Processing in Learning and Transfer by High and Low Imagers

Francis J. Di Vesta

Both Kuhlman (1960) and Stewart (1965) found that high imagers learned a paired-associate list most easily when the stimuli were pictures than when they were words. Furthermore, Stewart (1965) found that high imagers learned tasks in which pictorial stimuli were used more easily than did low imagers while low imagers learned lists in which verbal stimuli were used in fewer trials than did high imagers. These results imply that investigations of aptitude by treatment interactions (ATI) may be a useful method for investigating strategies employed by Ss during learning and recall.

The present study was first of all an attempt to replicate, in part, one of Stewart's experiments. Such replication appears to be especially justifiable in view of Cronbach and Snow's comments, as follows:

> . . . Progress toward the goal of identifying and understanding ATI has been slight. We have not examined every pertinent study but our survey has probed deeply enough to give us confidence that a truly exhaustive sample would not change the general picture as of this moment. There are no solidly established ATI relations even on a laboratory scale and no real sign of any hypothesis ready for application and development. There are intriguing findings here and there, none of which has been pursued through a sufficient series of replication, validity

generalization, and enhancement studies to make it impressive (1969, p. 193; italics ours).

Secondly, the major purpose of the present study was to investigate the effects of learning with pictorial and verbal stimuli on performance of a transfer task. It can be reasoned that pictorial representations tend to be more specific than a verbal stimulus for the same referent. Thus, while the pictorial stimulus, bird, will, under most circumstances, elicit the verbal mediator "bird" it will also be fairly specific in the sense that it cannot be a "template" for the abstraction of "all birds" unless it achieves symbolic status, as it might in abstract or expressionistic painting, at which point its equivalence to a verbal symbol can be assumed. In any diagrammatic representation there will be some restriction on what is perceived even though it is nothing more than the restriction that a class of "large birds" or of "small birds" is represented. On the other hand, the verbal stimulus "bird," or any other similar symbol for that matter, is more nearly representative of a highly generalized "template." In the sense that it represents a larger class of all experiences the S has had with birds, a verbal symbol should provide a broader base for transfer than the pictorial stimulus.

Based on the aforegoing rationale the present study extended Stewart's investigations through an experiment designed to examine the extent to which mediation processes differentially involved imaginal or verbal transformation of an experience. More specifically, the intent was to investigate whether one form of mediation takes precedence over another. It was hypothesized that if imagery was a dominant processing mechanism for storing and using concrete materials then transfer to pictorial representations (geometric representations or their verbal equivalents) should be facilitated when compared with transfer to verbal representations (particularly verbal subordinates). If verbal mediation was more advantageous than imaginal processing then, it was hypothesized, transfer to verbal representations would be more efficient than to pictorial representations.

It was also possible in this study to investigate a sequencing hypothesis. Thus, it was hypothesized that both imaginal and verbal mediation might be employed in processing and, if so, one possibly precedes the other. Accordingly, if imaginal precedes verbal mediation then the pictorial to verbal order of presentation was hypothesized to be more efficient than the verbal to pictorial order, and vice versa (Lockhart, 1969; Paivio, 1963; Yuille, Paivio, & Lambert, 1969). The hypotheses related to imagery, as an individual difference variable, extend the above hypotheses by implying that high imagers would be especially benefited by any treatment favoring imaginal processing while low imagers would be benefited by any treatment favoring verbal processing.

Method

General Design

Nouns which could also be graphically presented by basic geometric shapes served as stimuli during original learning. The S learned to associate either the verbal or pictorial representations of each of these nouns with a number from 2 through 9. The form of the stimuli (pictures or words) served as a between-subjects dimension. The Ss of each sex, which served as the other between-subjects dimension, were assigned in equal numbers to each experimental treatment.

The transfer task consisted of four lists in which the stimuli represented one of four relationships to the stimuli in the original learning task list. Thus, the lists for the transfer task were comprised of verbal subordinates, geometric representations, verbal equivalents of geometric representations, and identical representations but opposite modes of presentation of the object or word presented during initial learning. Each S received all four types of transfer stimuli which comprised a within-subjects dimension. Trials to criterion during original learning and number of correct responses to stimuli in each of the four transfer conditions served as the dependent measures. In addition to the mixed analyses of variance of these data, the effect of each of four aptitude variables on the number of correct responses was analyzed.

Stimulus Materials

Word (W) and picture (P) forms of eight generic nouns were paired with the digits 2 through 9 for the lists used in the original learning (OL) task. The transfer stimuli were either words (W) or pictures (P) in four relationships to the stimuli in the OL task. Thus, the transfer lists were: Condition SN, subordinates of the generic nouns presented during OL, e.g., the word or picture, dime, was used in the transfer task as a subordinate of the word coin used in the OL task; b) Condition GN, geometric representations or outline drawings of the generic noun presented during OL, e.g., a circle was used as a representational drawing of a coin; c) Condition GL, labels or word equivalents for the geometric figures used in Condition GN, e.g., the word "circle" rather than the figure might be presented in Condition GL; and Condition D, in which the representations in the transfer task were identical to those in the OL task but were presented in the opposite mode during OL, e.g., if a picture of a coin was presented during OL, the word "coin" would serve as a transfer stimulus, and vice versa. The stimuli used in all conditions are presented in Table 1.

Subjects

Undergraduate students at The Pennsylvania State University who were enrolled in an introductory course in educational psychology served as Ss. Although they were volunteers, the Ss received credit toward their final grade for participating in the experiment. All Ss had been previously administered a battery of tests designed to measure imagery and verbal ability. Assignment to learning conditions was randomized within sex of Ss. In all, 198 Ss, of which 94 were male, participated in this experiment.

Measures of Individual Differences

Measures of individual differences were obtained for Imagery-Ability, Verbal ability and Automatization. The Flags test (Thurstone & Jeffrey, 1959), Spatial Relations Test of the Differential Aptitude Test (Bennett, et al., 1963), and the Gottschaldt Figures Test as described by Thurstone (1944) were employed to measures Imagery-ability. The Stroop Color-Word test as described by Thurstone (1944) and the Automatization test (Broverman, Klaiber, Kobayaski, and Vogel (1968) were used to measure Automatization. A vocabulary test, reading test (both of which were locally constructed by compiling items from existing tests) and the College Examination Board's Scholastic Aptitude Test (1962-1963). The raw scores for tests associated with each factor were standardized for males and females separately. The resulting T-scores were summed and averaged to obtain an overall factor score. High- and

Table 1

Original and Transfer Stimuli for the

Learning and Transfer Conditions

	Original Learning	Transfer*	
Response	Generic Noun	Subordinate Noun	Word for Geometric Representation
2	Coin	Penny	Circle
3	Horn	Bugle	Cone
4	Cheese	Swiss	Wedge
5	Flower	Rose	Star
6	Snake	Python	Coil
7	Gem	Ruby	Hexagon
8	Tree	Spruce	Triangle
9	Drum	Snare	Cylinder

* Note -- The figures for the generic nouns were pictures of the objects listed under original learning. The figures for the geometric representations were pictures of the forms shown in the last column.

low-scorers were selected from the extremes of the resulting distributions of factor scores for men and women. High scorers had average T-scores greater than 55 and low scorers had average T-scores lower than 45.

Procedure

Upon arrival at the laboratory the S was randomly assigned to one or the other of the OL conditions. Instructions were provided as to the specific nature of the task, i.e., the S was either told that he would be presented word-number pairs or that he would be presented picture-number pairs. The study-test procedure with standard instructions was used.

During the study trial, stimulus pairs were rear-projected onto a translucent screen by a carousel projector at a 2-sec. rate. During the recall interval only the stimulus member was presented at a 2-sec. rate. The S's task in the recall phase was to respond with the number previously associated with it. Study-recall trials were administered until the S had identified all but one of the paired-digits correctly.

The transfer task was administered following a rest period of 2-min. The S was instructed that he would be presented 32 stimuli most of which he had not seen before but all of which had some relationship to the words learned during the OL task. He was instructed to respond with the same digits employed in OL and was told to base his response on possible relationships to the list he had just learned. Only one presentation of the transfer list was administered.

Results

Original Learning

The number of trials to criterion on the initial learning task were analyzed by a mixed analysis of variance. This analysis yielded *F* $(1,194) = 9.68$, $p < .01$ for the effect due to the kind of stimuli employed. The list based on words took more trials to learn ($\overline{X} = 3.50$) than did the list based on pictures ($\overline{X} = 2.64$). Thus, while the referent in each instance was the same for pictures and for words, pictorial depiction of stimuli clearly resulted in more rapid learning than symbolic presentation. These results suggest differential processing of stimuli presented by the two methods. The effects related to Sex of Subject or its interaction with the task variable were not significant ($p > .05$).

Transfer Performance

The number of correct responses for the transfer task were analyzed via a mixed analysis of variance with Sex of Subject and Mode of Presentation (W or P) during OL as the between-subjects variables and Kind of Stimulus (i.e., SN, GN, GL, and O) as the within-subjects variable. This analysis yielded *F* $(1,94) = 12.63$, $p < .001$, for the effect due to the Mode of Presentation during OL; *F* $(1,194) = 3.82$, $p = .05$ for the effect associated with Sex of Subject; *F* $(3,582) = 358.07$, $p < .001$ for the effect due to the Kind of Stimulus employed in the transfer task; and *F* $(3,582) = 7.20$, $p < .001$ for the interaction between Mode of Presentation during Original Learning and Kind of Stimulus employed on the transfer task. A summary of this analysis is presented in Table 2.

Table 2

Summary of Analysis of Variance

Of the Number of Correct Responses for All Subjects

Between Subjects	df	MS	F	
Sex (A)	1	27.01	3.82	$p = .05$
Presentation Mode:OL(B)	1	89.17	12.63	$p < .001$
A x B	1	2.25	0.32	
Error (b)	194	7.06		
Within Subjects				
Kind of Stimuli:Transfer (C)	3	478.04	358.07	$p < .001$
A x C	3	2.46	1.84	
B x C	3	9.62	7.20	$p < .001$
A x B x C	3	1.31	0.98	
Error (w)	582	1.34		

The women made more correct responses ($\overline{X} = 4.77$) on the transfer task than did the men ($\overline{X} = 4.39$). The means represented in the interaction between Mode of Presentation during OL and transfer are summarized in Table 3. The primary implication of the data in this table is that pictorial presentation of stimuli during original learning results in greater transfer than does verbal (symbolic) presentation for all transfer conditions; the difference, however, is least when the opposite mode or subordinate representation is employed in the transfer task. It was particularly disadvantageous for Ss to learn words on the first task and then transfer to a geometric representation whether that representation was in symbolic or pictorial form.

Individual Differences

The investigation of effects associated with individual differences was made by extending the design to include the high and low imagers as a third between-subjects factor. There were 10 Ss in each cell of this design.

This analysis yielded $F\ (1,72) = 13.20$, $p < .01$, for the effect associated with Imagery. The mean score for high-Imagers was higher ($\overline{X} = 4.90$) than that for low Imagers ($\overline{X} = 4.15$) on the transfer task. The hypothesized interaction between individual differences in visualization and treatments was not supported. None of the interactions was significant ($p > .05$).

The same analysis as that described immediately above was conducted by replacing the groups of Ss differing on Imagery scores with groups of Ss differing on Automatization scores. There were also 10 Ss in each group for this analysis, which is summarized in Table 4. The unique outcome of this analysis, compared to the earlier one, was that the third-order interaction was significant ($F\ [3,216] = 3.13$, $p < .05$).

Table 3

Mean Number of Correct Responses on the Transfer Task As Related to the Original Learning Task

Transfer Task Conditions	Presentation Mode: OL Words	Pictures	Difference
Opposite Modality	6.15	6.34	.19
Subordinate Representation (Words and Pictures Combined)	5.34	5.76	.42
Verbal Equivalent of Geometric Figure	2.55	3.51	.96
Geometric Figure Representation	2.90	4.03	1.13

Table 4

Summary Analysis of Variance Based on the Mean Number Of Correct Responses for High and Low Automatizers

Between Groups	MS	df	F	
Sex of Subject (A)	0.53	1	0.08	
Presentation Mode: OL(B)	52.00	1	8.04	< .01
Automatization (C)	22.58	1	3.49	< .10
A x B	1.65	1	0.26	
A x C	.01	1	0.00	
B x C	9.45	1	1.46	
A x B x C	.38	1	0.59	
Error (b)	.65	72		
Within				
Kind of Stimuli: Transfer (D)	215.14	3	180.24	< .01
A x D	1.99	3	1.66	
B x D	8.78	3	7.35	< .01
C x D	.57	3	0.48	
A x B x D	2.54	3	2.13	< .10
A x C x D	.38	3	0.31	
B x C x D	.84	3	0.70	
A x B x C x D	3.73	3	3.13	< .05
Error (w)	1.19	216		

The means for the groups represented in this interaction are summarized in Table 5. As in the previous analyses, all groups were found to function more effectively when pictures were employed in OL than when words were used. However, the main difference among groups to be identified in this table is the generally inferior performance of the women who were high Automatizers (i.e., those who take longer to perform the automatization tasks) relative to the performance of women who were low Automatizers (i.e., those who perform the automatization tasks in a shorter period of time) on all transfer tasks following original learning with word stimuli. The women low Automatizers who learned words on the original learning task averaged 1.10 more correct answers on the Transfer Task than did the high Automatizers. The average difference in performance on the transfer task between the same groups when pictures were used as stimuli during original learning was .03. Similar comparisons for males yielded averages of .65 (with words as stimuli in OL) and .47 (with pictures as stimuli) correct responses. If automatization can be considered as a measure of one kind of Imagery these results imply that the encoding by high-Imagers of verbal stimuli is more detrimental to their performance on new tasks than the encoding of pictorial stimuli. This difference holds for both women and men but less so for men.

Comparable analyses based on two other individual difference grouping, verbal and anxiety, were also made. The analysis of verbal groups yielded $\underline{F}$ (1,72) = 5.96, $\underline{p} < .05$ for differences in performance on the transfer task, associated with verbal ability. As would be expected, the group with the higher verbal ability made more correct responses ($\overline{X} = 4.81$) than did the groups with the lower verbal ability

53

Table 5

Mean Number of Correct Responses on the Transfer Task as Functions of Presentation Mode During Original Learning, Kind of Presentation During Transfer, Sex of Subject and Score on the Automatization Factor

Presentation Mode: OL	Transfer Task	Females High Automatizers	Females Low Automatizers	Difference	Males High Automatizers	Males Low Automatizers	Difference
Words	Opposite Modality	6.00	7.00	(1.00)	6.00	6.50	(0.50)
	Subordinate	5.40	6.50	(1.10)	4.60	5.10	(0.50)
	Geometric Figure	2.30	3.00	(0.70)	2.20	3.50	(1.30)
	Verbal Representation	1.50	3.10	(1.60)	2.40	2.70	(.30)
	Mean	3.80	4.90	(1.10)	3.80	4.45	(.65)
Pictures	Opposite Modality	6.20	6.10	(-0.10)	6.50	6.60	(0.10)
	Subordinate	6.00	6.00	(0.00)	5.30	6.70	(1.40)
	Geometric Figure	3.90	4.50	(0.60)	4.40	4.10	(-0.30)
	Verbal Representation	4.00	3.40	(-0.60)	3.00	3.70	(0.70)
	Mean	5.03	5.00	(- .03)	4.80	5.27	(.47)

Note.- The reader should keep in mind that a low automatizer does the task required on the automatization tests in a shorter period of time than does the high automatizer.

($\overline{X}$ = 4.15). In addition, this analysis yielded $F(1,72) = 6.19$, $p < .05$ for the effect associated with Sex of Subject; $F(3,216) = 141.14$, $p < .01$ for the effect due to Kind of Stimuli on the transfer task; and $F(3,216) = 3.29$, $p < .05$ for the interaction between Kind of Stimuli on the OL task and those on the transfer task. The major significance of the analyses described in this paragraph is to indicate only that where interactions of individual differences with treatments do occur, they are found with individual differences based on measures that appear to require some form of imaging. They do not occur with measures of cognitive processes or of personality (i.e., anxiety) behavioral tendencies.

It is interesting to note that the only personality or general intellectual ability measure to correlate with the transfer scores for female Ss (N = 104) was automatization. The correlations ranged from -.20 to -.25 ($p < .025$) for the four modes of presentation on the transfer task. However, the correlations between individual difference measures and transfer task measures for male subjects (N = 80) yielded ranges of .28 to .39 ($p < .01$) for the vocabulary score; .14 to .33 ($p < .05$) for the Gottschaldt Figures Test; and .23 to .35 ($p < .05$) for the SAT scores (both verbal and math). Thus, automatization was the only influential correlate with performance for women Ss while only imagery and verbal factors were influential correlates with performance for male Ss. These differences suggest a possible reason for differences between the results of the present study and those of Stewart (1965).

Discussion

Based on the results from the total group of Ss, it is clear that learning pairs of items with pictures as stimuli is easier, that is, it takes fewer trials, than learning with words as stimuli. This finding provides a direct replication of a parallel finding from Stewart's (1965) study.

However, Stewart also found strong aptitude (Imagers) by treatments interactions. Thus, high imagers were found to learn pairs with pictures as stimulus elements more rapidly than with words as stimulus elements. Low imagers were found not to differ when the two kinds of stimuli were used. In addition, no significant differences were found between the two groups in transferring from pictures to words or vice versa, although there was a tendency for the high imagers to be hindered to a greater extent than low imagers by going from words to pictures.

In this regard, the present study suggests that automatization may be the influential behavioral tendency in distinguishing Ss performance when they are required to transfer from words to pictures compared to transferring from pictures to words. While differences between high and low imagers did not interact with treatments in the present study it was found that high-automatizers (those who take longer to perform the task) were particularly handicapped in transferring from words to pictures. Note that this process is clearly correlated with the process involved in performance on the automatization tests. Thus, automatization is measured by facility in labeling pictures without interference from other contextual stimuli; that is, they must go from pictures to words quickly and accurately. For example, in the Stroop color-name test, the S is required to read the word printed on the card rapidly. He is able to do so to the extent that he does not experience interference from the color of the print which is always different from the color-word. Similarly, on the automatization task, the S must name, as rapidly and as accurately as possible, the three objects pictured a total of 100 times on the card. He is able to do so in this task to the extent that he does not experience interference from the memory traces of the pictures already read. Thus, the significant differences in the performance of high and low-automatizers on the transfer task appears to reflect the cognitive operations that distinguish the two groups. In this respect, the present study replicates the parallel finding from Stewart's study.

The findings regarding the interaction between the original learning and transfer tasks indicate that transferring from words to pictures is more difficult than to transfer from pictures to words. Again this finding replicates one obtained by Stewart (1965). Developmental variables are undoubtedly implicated in explaining this result. In this culture, the S's typical experience is to label an object or picture; the opposite requirement is rare indeed. Moreover, these results imply that Ss form a "percept" before providing a label for the learning experience. This strategy is assumed to be a more dominant one for high-automatizers than it is for low-automatizers. The debilitating effects on the performance of high automatizers were especially noted when the Ss transferred from words in the original learning task to geometric representations and labels for geometric representations in the transfer task. The Ss did have somewhat more difficulty in transferring from pictures to verbal or pictorial representations than they did to subordinate categories or to the opposite modalities. However, they made nearly twice as many correct responses on the transfer task with the representation stimuli when the picture to word sequence was employed than when the word to picture sequence was employed.

The present study suggests clear replications of treatment effects related to the use of picture versus words in presenting stimuli and of treatment effects related to the strategies employed by Ss in processing learning materials. The implication of the differences in the processing strategies of high and low automatizers is an intriguing one and appears worthy of further investigation. The results suggest the need for careful analysis of processes involved in measures of individual differences and even, perhaps, the employment of task-specific measures if aptitude by treatment interactions are to be found. Our experience with the tasks involved in this experiment suggested that ceiling effects were approached if not reached in its conduct. For example, the four lists in the transfer task, comprised of a total of 32 items, could probably have been answered correctly on the second trial. For this reason only one trial had been used. A more sensitive design, perhaps coupled with more sensitive measures such as latency, is clearly indicated.

References

Bennett, G. K., Seashore, H. G., & Wesman, A. G. Differential aptitude tests. (Grades 8-13 and adults.) New York: Psychological Corporation, 1963.

Broverman, D. M., Klaiber, E. L., Kobayashi, Y., & Vogel, W. Roles of activation and inhibition in sex differences in cognitive abilities. Psychological Review, 1968, 75, 23-50.

College Entrance Examination Board Scholastic Aptitude Test. Princeton, N. J.: Educational Testing Service, 1962-1963.

Kuhlman, C. K. Visual imagery in children. Radcliffe College, Unpublished doctoral thesis, 1960.

Lockhart, R. S. Retrieval asymmetry in the recall of adjectives and nouns. Journal of Experimental Psychology, 1969, 79, 12-17.

Paivio, A. Learning of adjective-noun paired-associates as a function of adjective-noun word order and noun abstractness. Canadian Journal of Psychology, 1963, 17, 370-379.

Stewart, J. C. An experimental investigation of imagery. Unpublished doctoral dissertation, The University of Toronto, 1965.

Thurstone, L. L. A factorial study of perception. Chicago: University of Chicago Press, 1944.

Thurstone, L. L., & Jeffrey, T. G. Space thinking (Flags). Chicago: Education-Industry Service, 1959.

Yuille, J. C., Paivio, A., & Lambert, W. E. Noun and adjective imagery and order in paired-associate learning by French and English subjects. Canadian Journal of Psychology, 1969, 23, 259-466.

The Effects of Rated Vividness and Imagery of Learning Materials On Learning and Recall of High- and Low-Imagers

Francis J. Di Vesta, G. Susan Gray, Gary M. Ingersoll,
Steven Ross and Phyllis Sunshine

The learning strategy that differentiates high and low imagers was examined, albeit indirectly, in this study. Previous studies permit the conclusion that recall of words is affected by the qualities of vividness (Tulving, McNulty and Ozier, 1965) or imagery (Paivio, 1970). Furthermore, the degree of subjective organization has been found to vary according to the vividness of the list. For example, Tulving, McNulty and Ozier (1965) comment, "But the fact that the recall of more vivid words was ... organized by subjects to a greater extent thatn that of less vivid words is compatible with the hypothesis that vividness or picturability is an important component of meaning of words that affects the ease with which words can be grouped into higher-order memory units" (p. 250). However, these authors caution that such explanations are often inadequate since experiments in which only levels of vividness are manipulated deal only with correlations between stimulus characteristics and learning. Thus, for example, in the Tulving, et al. study, the effect might have been due to concreteness (or abstractness) or to associative relationships among the words in a given list rather than to "picturability" or other imaginal properties of the stimuli.

In order to examine whether some organizational process related to imagery was employed by learners in such tasks as those used by Tulving, et al. (1965), Stewart (1965) examined the interaction between ability to use imagery as measured by the Flags (Thurstone & Jeffrey, 1959) and Spatial Relations (Bennett, Seashore, & Wesman, 1963) tests and the characteristics of the material learned. Exactly the same lists described by Tulving, et al. (1965) were used. As in the earlier study, the concrete (vivid) words were recalled more readily than abstract words. More interesting was the finding by Stewart (1965) that with these same lists the material was increasingly organized by high imagers as the vividness ratings of the words in the lists were increased while there were no differences in the degree of organization among lists by the low-imagers.

Stewart's results provide one basis for making inferences about the kind of processes that are employed by learners when approaching different learning tasks. They suggest too, that some treatments are more effective for learners who are imagers than for learners who are non-imagers. Since vividness (or picturability) implies an imaginal characteristic of symbolic stimuli, it would appear that imaginal transformations of stimuli (that is, words into "pictures") are as relevant as other cognitive processes in organizing materials for storage in memory.

Because of its implications for understanding strategies for learning and recall as well as for investigations of aptitude by treatment interactions, the present study was conducted to determine the replicability of Stewart's study. Furthermore, after a careful review of the literature on aptitude by treatment interactions, Cronbach and

Snow (1969) conclude that investigators have had difficulty in obtaining replications of such interactions, thus providing an important justification for attempting a replication of Stewart's study. The present study extends the earlier one by incorporating not only the original lists of words but also a set of lists based on the norms published by Paivio, Yuille, and Madigan (1968). As in the earlier study, the purpose of the present one was to investigate the hypothesis that the learning and recall of high-imagers is significantly greater when learning concrete words than when learning abstract words but learning of low-imagers is not differentially related to the concreteness or abstractness of the stimuli.

Method

Design

The overall procedure consisted of Ss learning lists of words that were high, medium, and low in concreteness (i.e., either rated vividness or rated imagery). The lists were presented in counterbalanced order. Each S was presented four trials of each list via the study-free recall procedure. The sequence of trials was different for each S learning a given list. The important features of the design implied a 2 x 3 mixed analysis of variance in which there were two levels of the between-subjects factor (vividness or imagery lists) and three levels of the within-subjects factor (high, medium, or low concreteness). Some of the analyses of variance included two levels of individual differences (either high and low imagers or male and female Ss) and order of presentation of the lists as between subjects factors. Correlations were computed for relationships between imagery, verbal ability and clustering scores.

Subjects

A total of 219 students enrolled in an introductory educational psychology course served as Ss in this study. Of these 104 were males and 115 were females. All Ss participated voluntarily but received credit toward their final grade for such participation.

Materials

Two sets of three word lists were employed: one set was based on rated vividness (V) and the other on rated imagery (I). The three lists (high, medium, and low) of 16 words within each set differed in the degree of rated-vividness or rated-imagery of their constituent words but were very nearly the same in terms of meaningfulness and Thorndike-Lorge (1944) frequency.

The V set of lists was identical to that described by Tulving, et al. (1965). In obtaining their values for vividness, they defined vividness as "the ease with which something could be pictured in the mind." Ratings were obtained on a 7-point scale with 1 corresponding to no image and 7 to extreme vividness. Meaningfulness of a word was obtained by ratings on a scale of 1 (corresponding to meaningless) to 7 (corresponding to extreme meaningfulness). The three lists in the V set are presented in Table 1.

The I set of lists was prepared from the Paivio, Yuille, and Madigan (1968) list of 1000 words rated for imagery and meaningfulness. The three lists in the I set were constructed in essentially the same manner as described for the V set. In the Paivio, et al. (1968) norms imagery was defined as the ease or difficulty with which a mental image

Table 1

The Experimental Lists of Words Based on Vividness Ratings*

	High Vivid (Concrete)			Medium Vivid			Low Vivid (Abstract)	
Word	V	M**	Word	V	M**	Word	V	M**
Apron	5.17	4.00	Abode	4.13	4.06	Buyer	2.94	3.93
Balloon	5.63	4.03	Bucket	4.50	3.88	Crisis	2.10	5.23
Bunny	5.23	3.93	Builder	4.00	4.28	Entry	2.08	3.71
Butler	5.79	3.79	Cargo	4.57	4.03	Founder	1.85	3.92
Cabbage	5.32	4.06	Fiber	3.96	3.50	Output	1.50	3.77
Camel	5.79	3.88	Hamlet	4.83	4.03	Patron	2.70	3.50
Chorus	5.63	4.43	Handful	4.36	3.43	Renown	1.33	3.94
Cigar	6.53	4.41	Madame	4.61	4.27	Routine	1.28	4.64
Circus	6.08	4.54	Pebble	4.76	4.15	Rover	2.11	3.30
Comet	5.81	3.96	Porter	4.64	3.79	Rumour	2.20	4.57
Granny	5.00	4.17	Pudding	4.61	4.05	Session	2.86	3.83
Jungle	6.00	4.36	Summit	4.77	4.46	Surplus	1.83	4.43
Lantern	5.43	4.32	Thicket	4.54	3.75	Tariff	1.50	3.61
Rainbow	6.13	4.94	Trainer	4.25	4.19	Topic	1.50	4.03
Runner	5.64	4.61	Veteran	4.07	4.40	Treason	1.50	4.70
Satin	5.42	3.62	Voter	4.32	5.15	Vigour	2.70	4.44
Mean	5.66	4.19	Mean	4.43	4.09	Mean	2.00	4.10

* From Table II, page 245, in Tulving, McNulty and Ozier (1965).

** Meaningfulness based on ratings.

was aroused by the word. Meaning was measured by Noble's (1952) production method. The lists of words in the I set with their associated imagery and meaningfulness values are presented in Table 2.

There were prepared 16 randomized trials of the words comprising each of the three lists within both the V and I sets, following the procedure described by Stewart (1965); that is, a given word did not occur in the same serial position, was not preceded by the same word, and was not followed by the same word more than once in all trials, eight of which were mirror images of the others. Each set of the 16 trials for a given list was placed on a single memory drum tape.

Procedure

The stimuli were presented to the Ss on a Stowe Memory Drum at a rate of one word per second. A given S was presented each of the three lists within a set: that is, he was presented all three versions, high (H), medium (M), and low (L), of either the I set or of the V set. The order of presentation of the lists within a set was randomized from one S to the next with the restriction that each of the possible combinations of list orders (LMH, LHM, MHL, MLH, HLM and HML) was equally represented at the completion of the study.

The Ss were administered four trials of a given list, the first of which was randomly selected from the 16 on the memory drum tape. The study-free-recall procedure, with standard instructions, was used. During the study period the S read aloud each of the 16 words in the list as they were presented. The end of the list was signalled by a dotted line after which the S was to write down all the words he could remember and in any order he chose. The free recall period was 90-secs.

Table 2

The Experimental Lists of Words Based on Imagery Ratings

Word	High Imagery (Concrete) I	M*	Word	Medium Imagery I	M	Word	Low Imagery (Abstract) I	M
Acrobat	6.53	5.67	Anger	4.87	5.83	Belief	2.73	5.24
Alcohol	6.47	6.08	Assault	4.80	5.56	Crisis	3.43	5.44
Barrel	6.57	6.16	Barnacle	4.50	5.69	Edition	3.40	5.88
Bouquet	6.77	5.76	Belfry	4.43	6.25	Evidence	3.23	5.[illegible]
Candy	6.63	6.39	Blessing	4.43	6.19	Gender	2.90	5.41
Cigar	6.80	6.22	Chaos	4.57	5.88	Intellect	2.93	5.56
Factory	6.43	6.00	Clearness	4.77	5.64	Irony	2.83	5.24
Headlight	6.43	6.32	Comedy	4.90	6.08	Magnitude	2.50	5.68
Hotel	6.40	5.96	Humor	4.57	5.72	Malady	3.37	6.00
Hurdle	6.33	5.92	Mirage	4.97	5.63	Mastery	2.77	5.46
Jelly	6.40	6.00	Portal	5.10	5.63	Miracle	3.33	5.60
Piano	6.70	6.48	Reflex	4.73	5.88	Moral	3.17	6.44
Scarlet	6.37	5.80	Revolt	5.07	5.60	Origin	2.30	5.32
Slipper	6.47	6.04	Vacuum	4.77	5.94	Perjury	3.37	5.92
Steamer	6.53	6.32	Vapour	4.80	5.76	Satire	3.37	5.64
Tweezers	6.57	5.80	Victory	4.93	6.12	Welfare	3.17	6.16
Mean	6.52	6.05	Mean	4.76	5.84	Mean	3.05	5.70

* Meaningfulness based on Noble's (1952) production method.

long. This procedure was followed for all four presentations of a list following which there was a 2-min. rest period. The S was then administered the next series of items, either H, M, or L, depending on the condition to which he had been randomly assigned. The procedure was repeated until all three lists within a set had been presented.

Tests*

The following tests, fully described in another report (Di Vesta, Ingersoll, & Sunshine, 1971, in press) were administered to all Ss: the Space Thinking (Flags) test, the Spatial Relations test from the Differential Aptitude Test Battery, the Gottschaldt Figures test, the Stroop Color-Name test, the Automatization test, the Scholastic Aptitude (Math and Verbal) test, a vocabulary test, the Remote Associates test, a Reading Comprehension test, the Achievement Anxiety Scale and the Dogmatism Scale. The data were factor analyzed by the principal components method for the initial factorization. When the factors were rotated via the Varimax routine, three factors of relevance to this study were extracted: Verbal (as represented by the Scholastic Aptitude Test Verbal score), Imagery (as represented by the Spatial Relations test), and Automatization (as represented by the Stroop Color-Name). Raw scores for each test were standardized. The standardized scores for each test saturated on a given factor were summed to obtain a factor score. The Imagery factor score was comprised of the Flags, Spatial Relations, and Gottschaldt Figures Tests. The Ss who attained the highest 25 scores were classified as low Imagers. The distributions for men and women were considered separately. The same procedure was employed for high and low Automatizers based on the Stroop Color-Name

*The tests are described fully in the report entitled "A Factor Analysis of Imagery Tests" by Di Vesta, Ingersoll, & Sunshine, presented elsewhere in this report.

test and Broverman's Automatization Test; and for high and low Verbal Ability based on the Scholastic Aptitude Test: Verbal Score, a Vocabulary test, a Reading Comprehension test, and the Remote Associates Test.

Results

Overall Analyses of Number of Correct Responses

The number of correct responses on the recall trials were analyzed initially by an overall mixed analysis of variance in which measures of dispositional traits were ignored. This analysis was conducted to examine and determine the gross effects of manipulated variables and to aid in decisions about pooling of data for subsequent analyses. There were three between subjects variables: Sex of Subject, Kind of Lists (Vividness or Imagery), and six Orders of Presentation of lists. The within subject variables were: Levels of Concreteness (high, medium and low) and four Trials.

This analysis yielded F (1,168) = 46.89, $p < .001$ for the effect due to Sex of Subject; F (2,336) = 64.74, $p < .001$ for the effect due to Levels of Concreteness; F (3,504) = 726.64, $p < .001$ for the effect due to Trials; and F (3,504) = 18.44, $p < .001$ for the effect due to the interaction of Sex of Subject and Trials. None of the other interactions were significant ($p > .05$). The complete summary table for this analysis is displayed in Table 3.

These results indicated that women Ss averaged more correct responses ($\overline{X}$ = 10.68) per trial than did the men ($\overline{X}$ = 7.30). The Ss averaged more correct responses per trial on the concrete list ($\overline{X}$ = 9.47), than on the medium concreteness ($\overline{X}$ = 9.15) or the abstract ($\overline{X}$ = 8.35) lists. Average numbers of correct responses were 6.05 for the first,

Table 3

Summary of Overall Analysis of Variance of Correct Responses

Between Subjects	MS	df	F	p
Sex (A)	6444.14	1	46.89	<.001
Lists (B)	144.50	1	1.04	
Orders (C)	8.94	5	0.06	
A x B	6.15	1	0.04	
A x C	12.99	5	0.09	
B x C	21.91	5	0.16	
A x B x C	8.67	5	0.06	
Error (b)	139.55	168		
Within Subjects				
Levels (D)	257.05	2	64.74	<.001
A x D	2.09	2	0.53	
B x D	8.54	2	2.15	
C x D	6.93	10	1.75	<.10
A x B x D	6.93	2	1.75	
A x C x D	2.28	10	0.56	
B x C x D	3.20	10	0.81	
A x B x C x D	4.19	10	1.06	
Error (w)	3.97	336		
Trials (E)	2666.63	3	726.64	<.001
A x E	67.66	3	18.44	<.001
B x E	1.99	3	0.54	
C x E	1.85	15	0.50	
A x B x E	.74	3	0.20	
A x C x E	1.26	15	0.34	
B x C x E	2.23	15	0.61	
A x B x C x E	2.60	15	0.71	
Error (w)	3.67	504		
D x E	1.18	6	0.71	
A x D x E	.96	6	0.58	
B x D x E	3.15	6	1.89	<.10
C x D x E	2.17	30	1.30	
A x B x D x E	1.01	6	0.60	
A x C x D x E	1.34	30	0.81	
B x C x D x E	1.74	30	1.05	
A x B x C x D x E	1.13	30	0.68	
Error (w)	1.67	1008		

8.79 for the second, 10.16 for the third, and 10.95 for the fourth trials. The Sex by Trials interaction indicated that the women Ss learned the list more rapidly than men. Additional descriptions of the sex differences in learning these tasks will be provided in the analyses of individual differences.

The findings from this analysis are in substantial agreement with those obtained by Stewart (1965) and Tulving, et al. (1965). However, procedural differences between the ones used in those investigations and the present study should be noted as follows: In the present study four trials were used instead of eight, both men and women were used as Ss rather than only women, and lists varied in terms of imagery as well as vividness were employed. Despite these differences the finding that concrete lists are learned more readily than abstract lists was clearly replicated and performance over trials closely approximated the performance of Tulving, et al.'s (1965) Ss at the end of four trials. In addition, it was found that women learn the task more readily than men and that the Tulving, et al. (1965) lists produce essentially the same results as the lists based on the norms published by Paivio, et al. (1968). Unlike the earlier studies, practice effects or learning-to-learn were not observed in the present study. The reason for this difference may be due partly to the confounding of Lists and Orders and partly to the use of fewer trials than in the earlier studies. However, it should be noted that Stewart, who also confounded orders and lists found only a very small effect due to practice; her Ss averaged, over all trials: $\overline{X}$ = 13.02 for the first list learned, $\overline{X}$ = 13.44 words for the second list; and $\overline{X}$ = 13.56 for the third list.

Individual Differences in Imagery Related to Performance

The subsequent analyses of the number of correct responses were made by mixed analyses of variance with individual differences in Imagery (high and low) as the between variable and Trials and Levels of Concreteness (H, M, and L) as the within variables. Since the main effect of Sex as a factor and the interaction of Sex by Trials was significant in the initial analyses, separate analyses were made for men and women. Conversely, because there were no significant differences due to Kind of Lists in the previous analyses, the data for Ss administered the V or I lists were pooled for the present analyses. (See Tables 4 and 5.)

The analysis of variance of data for the men yielded no significant differences ($p > .10$) for the main effect of Imagery; F (2,96) = 16.85, $p < .001$, for the effect due to Levels of Concreteness; and F (3,144) = 431.57, $p < .001$ for the effect related to Trials. The effect of primary concern in this study, however, is that due to the interaction between Imagers and Level of Concreteness which yielded F (2,96) = 2.47, $p < .10 > .05$. The data comprising this interaction are presented for men and women Ss, separately, in Figure 1, and for the pooled groups in Figure 2.

The analysis of data for women yielded F (1,48) = 3.11, $p < .10$ for the effect due to Imagery levels. The main effects due to Levels of Concreteness and to Trials were significant ($p < .001$) as in the previous analyses. The effect due to the interaction of Imagery by Trials yielded F (3,144) = 4.71, $p < .01$, the data for which are presented in Figure 3. This ordinal interaction is identical to the one obtained by Stewart (1965) in all essential respects. None of the other interactions were significant ($p > .10$) in this analysis.

Table 4

Summary Table for Analysis of Variance of Individual Differences: Female

		Individual Difference Variable							
		Imagery		Automatization		Verbal		Anxiety	
	df	MS	F	MS	F	MS	F	MS	F
Between Groups									
Individual Difference (A)	1	65.34	3.11[a]	56.43	2.41	5.80	.262	17.68	0.85
Error (b)	48	21.03		23.39		22.18		20.78	
Within Subjects									
Levels (B)	2	45.50	12.62[c]	51.13	12.50[c]	51.81	12.25[c]	25.90	7.00
A x B	2	4.34	1.13	12.05	2.95[a]	6.59	1.56	5.13	1.39
Error (w_1)	96	3.84		4.09		4.23		3.70	
Trials (C)	3	919.23	569.31[c]	958.92	376.47[c]	1022.64	483.26[c]	060.04	459.30[c]
A x C	3	7.60	4.71[b]	1.32	0.52	3.09	1.46	.99	0.47
Error (w_2)	144	1.61		2.55		2.12		2.11	
B x C	6	1.45	0.83	1.40	0.77	1.94	1.05	2.55	1.46
A x B x C	6	3.01	1.78	1.35	0.74	1.83	1.00	1.05	0.60
Error (w_3)(	288	1.74		1.83		1.85		1.74	

[a] $p < .10 > .05$

[b] $p < .01$

[c] $p < .001$

Table 5

Summary Table for Analysis of Variance of Correct Responses: Male

		Individual Difference Variable							
		Imagery		Automatization		Verbal		Anxiety	
	df	MS	F	MS	F	MS	F	MS	F
Between Groups									
Individual Difference (A)	1	8.88	.445	135.37	6.95[a]	24.40	1.15	105.8	5.71[a]
Error (b)	48	19.94		19.47		21.25		18.52	
Within Subjects									
Levels (B)	2	73.75	16.85[c]	49.01	10.20[c]	113.29	26.12[c]	74.48	18.72[c]
A x B	2	10.80	2.47[a]	2.86	0.59	8.21	1.89	4.58	1.15
Error (w_1)	96	4.37		4.80		4.34		3.98	
Trials (C)	3	817.69	431.57[c]	846.92	495.58[c]	836.61	399.71[c]	848.32	429.43[c]
A x C	3	.006	0.03	3.19	1.87	1.40	.67	1.36	0.69
Error (w_2)	144	1.89		1.71		2.09		1.98	
B x C	6	2.18	1.17	1.21	.62	2.44	1.12	2.63	1.62
A x B x C	6	1.77	0.95	1.70	.87	1.40	.64	2.34	1.44
Error (w_3)	288	1.87		2.00		2.19		1.62	

[a] $p < .10 > .05$

[b] $p < .05$

[c] $p < .001$

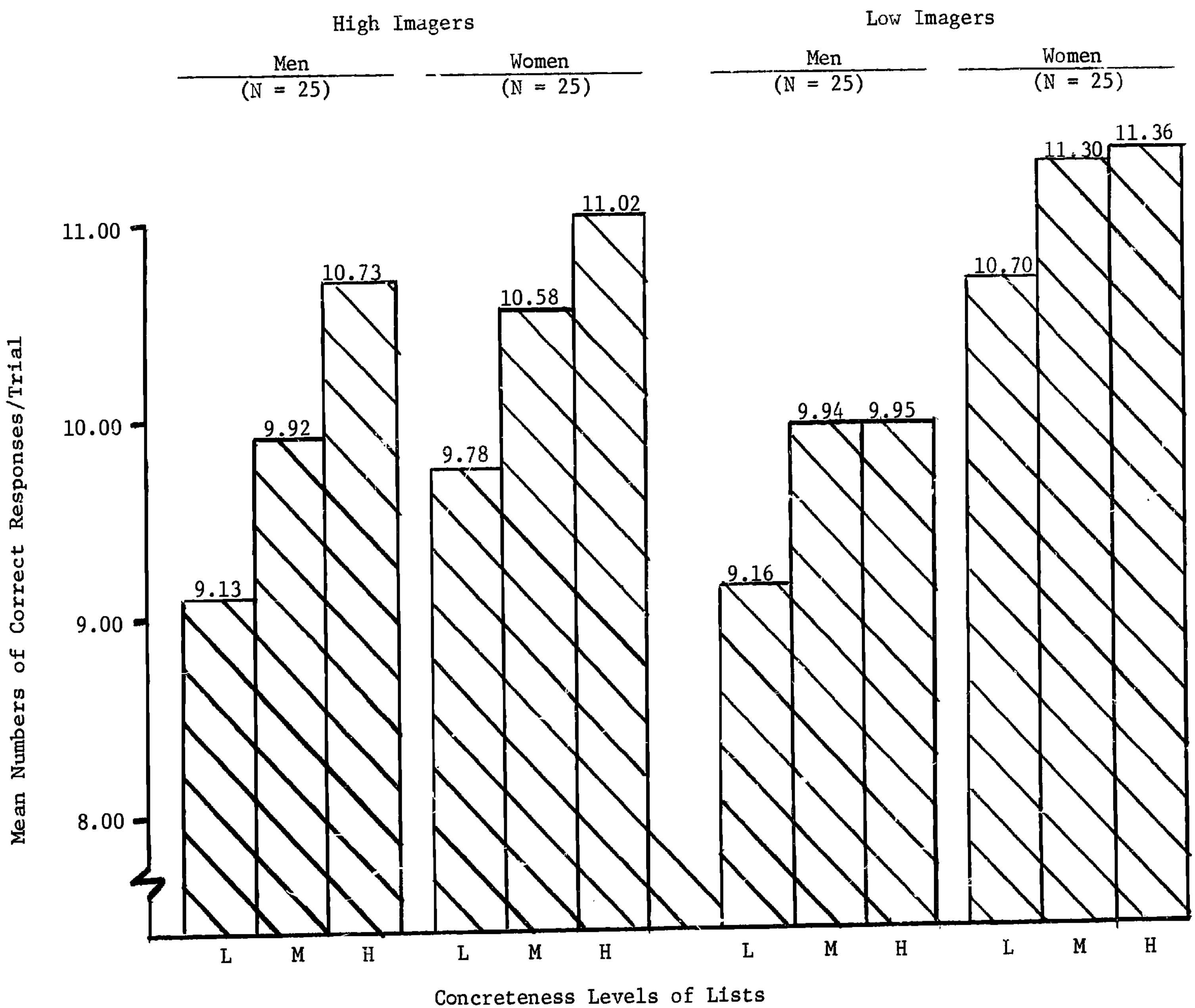

Figure 1. Mean numbers of correct responses made by men and women high and low imagers on lists comprised of low (L), medium (M), and high (H) concreteness words.

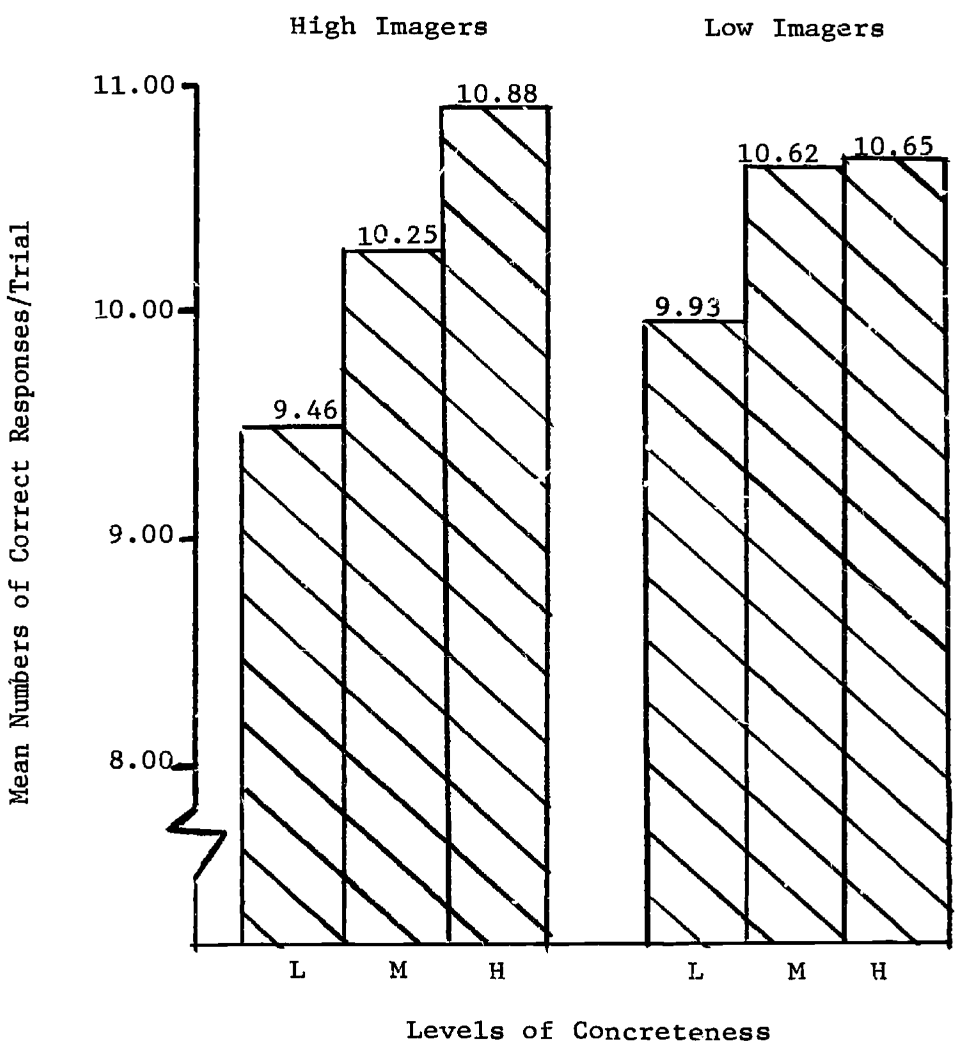

Figure 2. Mean numbers of correct responses on lists of low (L), medium (M), and high (H) concreteness words by high and low imagers.

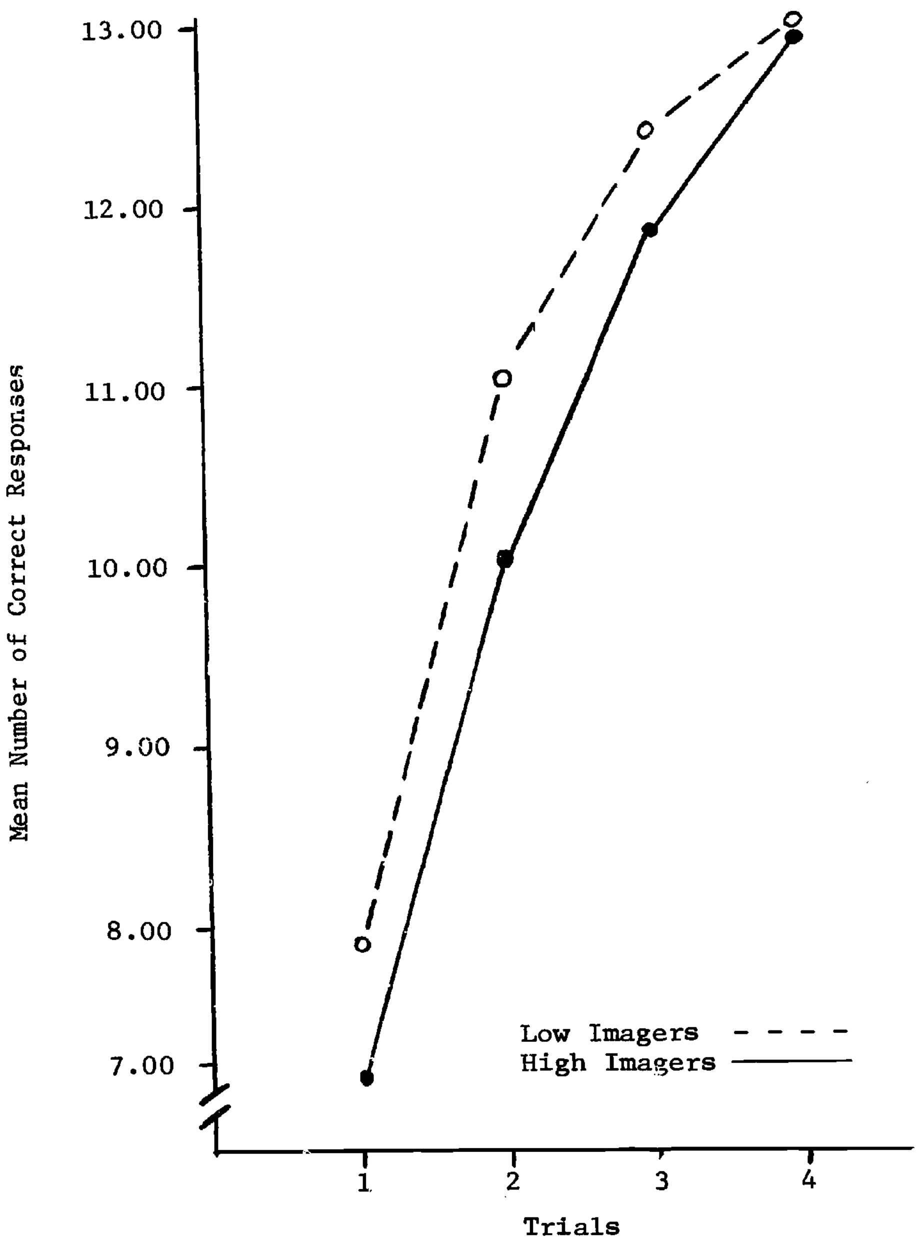

Figure 3. Mean number of correct responses for high- and low-imagers over trials for women subjects only.

The differences between Imagery levels, though not significant is interesting because the low imagers averaged more correct responses ($\overline{X}$ = 11.12) than did the high imagers ($\overline{X}$ = 10.46) over all trials. The differential performance of these two groups is more analytically reflected in the Imagery by Trials interaction. In this interaction (p < .01) the high imagers perform at a much lower level ($\overline{X}$ = 6.85) than the low imagers ($\overline{X}$ = 7.92) on the first trial but they perform at about the same level ($\overline{X}$ = 12.94) as the low imagers ($\overline{X}$ = 13.04) on the fourth trial. Although the interaction between Imagers and Levels of Concreteness for the male Ss was not significant (p > .10), their performance was very much like the women's performance and so have been presented in Figure 1 for purposes of comparison.

In order to provide a more direct comparison with the results from Stewart's (1965) study, another analysis was conducted identical to that described immediately above except that the high and low imagers were selected on the basis of only the Flags and Spatial Relations tests. The raw scores from each test were standardized and the two standardized scores for each S were then averaged. The women with the top 25 ranks on these scores comprised the group of high imagers and those with the bottom 25 ranks comprised the group of low imagers.

This analysis yielded significant differences (p < .001) for the main effect due to Levels of Concreteness and to Trials. The effect due to the interaction between Imagery and Levels of Concreteness yielded F (2,96) = 2.41, p < .10 and that due to the interaction of Imagery and Trials yielded F (3,144) = 3.27, p < .05. These effects were essentially the same as displayed in Figure 1, 2, and 3.

The Imagery by Levels of Concreteness interaction, in this analysis where only the Flags and Spatial Relations tests were used to identify high and low imagers, indicated that women high Imagers obtained an average number of 9.85, 10.98 and 11.26 correct responses per trial for the low, medium, and high concreteness lists, respectively, while the women low imagers obtained an average of 10.40, 11.10 and 10.96 correct responses per trial for the low, medium, and high concreteness lists, respectively. As can be noted in Figure 1, the previous analyses in which factor scores were used yielded identical trends, though the interaction was not significant. As with data obtained from men Ss, these findings imply that high imagers benefit by increases in vividness or concreteness of the stimulus material where a strategy involving imaginal processes are employed. When concrete stimulus material is used, high imagers perform as adequately as the low imagers. The latter, on the other hand, perform with relatively equal efficiency on all tasks, though slightly less so on the tasks involving concrete materials; that is, the low imagers strategy for learning was relatively unaffected (unrelated) by the vividness of the task materials.

The performance differences between high and low imagers take on increased importance when compared with similar analyses based on other individual differences. Thus, identical analyses to those for imagery were made with the individual difference variables based on automatization, anxiety, and verbal scores. Significant differences ($p < .05$) were found for the main effects of automatization and anxiety based on the analysis of data for male Ss only. None of the other main effects (that is, the verbal factor for data based on the responses of the men, and automatization, anxiety, and verbal individual differences for data based on the responses of the women) were significant ($p > .05$). However, the main effects associated with Trials and Levels of Concreteness were, as might be expected, significant ($p < .01$) in every analysis. Among all possible interactions in these analyses only that between Automatization and Levels of Concreteness, based on the data for women Ss approached significance ($p > .05 < .10$). This interaction was very similar to that reported above for imagery differences and seems to be reasonable since the measures of Automatization were comprised of tasks involving visual materials.

Effect of Imagery-ability on Organization During Recall*

An analysis was made of the free-recall data on Trials 1 through 4 inclusive of the number of intertrial repetitions (ITR) and the number of correct responses common to trial N and trial N + 1 (ITC). The ITR's were computed following the procedure described by Bousfield and Bousfield (1966) and corrected by subtracting the expected ITR's from the observed ITR's. The ITC's were computed simply by counting the number of correct responses on trial N that were also on Trial N + 1. The number of new responses on trial N + 1 compared to trial N (ITN) were also computed by simple count. Finally, a sequential consistency (SC) score, which is an ITR score based on ratios, was computed according to a description by Fagan (1968).

These data were analyzed initially for possible differences due to list. Since there wre no significant list differences for either measure the lists were ignored in all subsequent analyses. Nevertheless, the two lists were equally represented in all conditions of the analyses described below. Although scores from all measures were analyzed, only

* The authors are indebted to Professor Susan Rosner at the University of Iowa, for the loan of her program for computing the several clustering scores indicated in the heading of Table 6 and Table 7.

the data for the ITR and ITC yielded trends of interest. Accordingly, these are the only analyses summarized here though the data for all scores are summarized in the accompanying tables of correlations to be presented below.

The scores derived from the two measures were analyzed in separate, mixed analyses of variance in which the between-subjects variables were Sex of subject and Imagery-ability. There were 25 Ss of each sex represented in the high imagery-ability and low imagery-ability groups, respectively. The within-subjects variables were Concreteness (rated imagery) of lists, and Trials compared (i.e., Trial 1 vs. Trial 2, Trial 2 vs. Trial 3, and Trial 3 vs. Trial 4).

The analysis of the ITR data yielded $F\ (2,192) = 5.90$, $p = .003$ for the effect due to Concreteness of list, and $F\ (2,192) = 9.34$, $p < .001$ for the effect due to Trials. The analysis of the ITC data yielded $F\ (1,96) = 11.08$, $p < .001$ for the effect related to Sex of subject; $F\ (2,192) = 23.26$, $p < .001$ for the effect due to Concreteness of List; $F\ (2,192) = 767.31$, $p < .01$ for the effect related to comparison among Trials; and $F\ (2,192) = 4.67$, $p < .01$ for the effect due to the interaction of Trials by Sex of Subject. None of the other main effects of interactions were significant in either analyses. In general, these data indicate, especially for the ITC scores, that women Ss organized the words ($\overline{X} = 8.36$) more than did men ($\overline{X} = 7.34$); that there is more organization of lists with high concrete words ($\overline{X} = 8.45$) than of lists with medium concreteness ($\overline{X} = 7.97$), or low concreteness ($\overline{X} = 7.14$); and there were fewer responses in common to Trials 1 and 2 ($\overline{X} = 5.65$), than there were on Trials 2 and 3 ($\overline{X} = 8.20$) or on Trials 3 and 4 ($\overline{X} = 9.70$).

Concreteness clearly influences the ease with which a list is subjectively organized but the analyses failed to yield a significant interaction with imagery-ability except for a marginally significant ($p < .10$) interaction between imagery-ability and concreteness of lists when the ITC score was used as the dependent variable (see Figure 4). Accordingly, it was decided to compute correlations between imagery-ability scores and the several clustering measures separately for each of the lists representing different levels of concreteness. The entire sample of Ss (N = 219) was used in this analysis, the results of which are summarized in Table 6. The correlations computed separately for men and women Ss are also shown in that table.

As can be seen, there were significant correlations between imagery-ability and ITR scores for recall on the last two trials of the low-concreteness list ($r = -.14$, $p < .05$) and the first two trials of the high-concreteness list ($r = -.15$, $p < .05$). Although both are low correlations, they indicate that low-imagers achieve higher ITR's than do high imagers.

It is important, of course, to determine whether the relationships described above are confounded with some other ability. Unfortunately, it would be impossible to ferret out all such possibilities, but the most important one appeared to be the relationship between organization during recall and verbal (or general intelligence) ability. Accordingly, parallel correlations to those shown in Table 6 were computed by employing scores based on the tests comprising the verbal factor. These correlations are summarized in Table 7. None of these correlations was significant at the .05 level confidence, for data based on the total group the relationship between verbal ability and organization approached

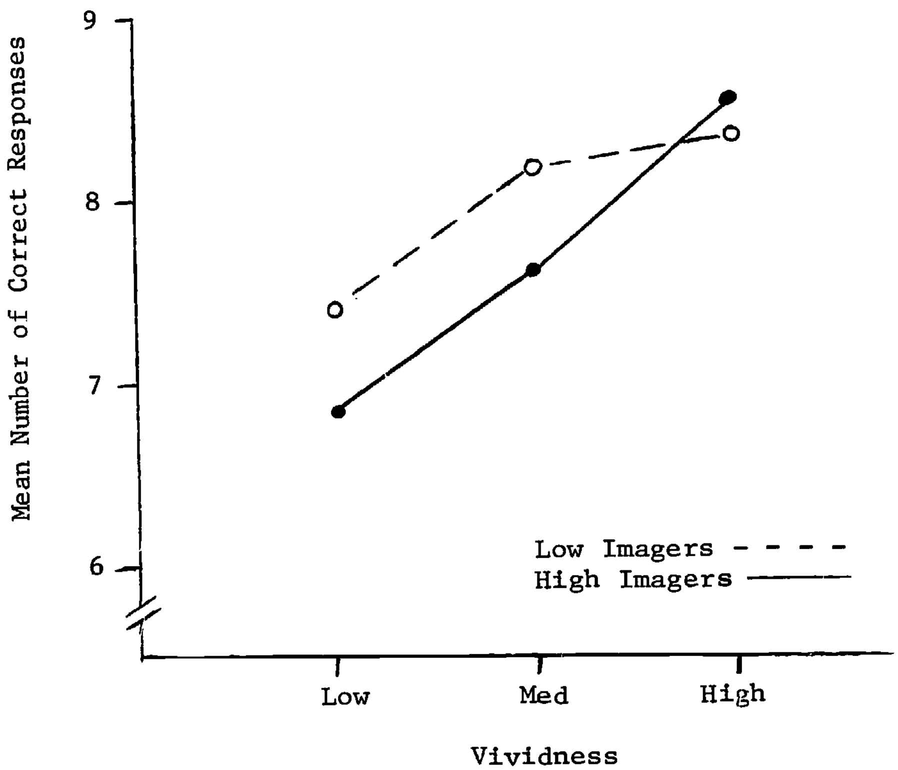

Figure 4. Mean number of correct words common to adjacent trials as a function of imagery-ability and list vividness.

Table 6

Correlations Between Measures of Organization in Recall and Imagery-Ability for Males and Females

List Imagery and Sex of Subject	Intertrial Repetitions (ITR)			Sequential Consistency			Intertrial Correct Responses (ITC)			Intertrial New Responses (ITN)		
	1-2	2-3	3-4	1-2	2-3	3-4	1-2	2-3	3-4	1-2	2-3	3-4
<u>High Imagery List</u>												
Males	-.05	-.06	-.01	-.10	-.06	-.04	-.04	-.17*	-.16	-.10	-.01	-.02
Females	-.24***	-.06	.05	-.19**	-.13	.02	-.18*	-.08	-.06	.04	.12	.00
Total Group	-.14*	-.06	-.03	-.14**	-.09	-.01	-.08	-.02	.03	.05	.06	-.01
<u>Medium Imagery List</u>												
Males	-.04	.03	.08	.11	.01	.09	-.18*	-.09	-.03	.15	.01	.07
Females	-.09	-.08	-.03	-.06	-.04	-.02	-.28***	-.14	-.11	-.04	.06	.23**
Total Group	-.07	-.03	.01	.02	-.02	.03	-.23**	-.12*	-.08	-.08	.04	.16**
<u>Low Imagery List</u>												
Males	-.03	-.02	-.05	-.01	.03	-.06	-.03	-.07	-.00	-.01	-.04	-.03
Females	-.13	-.09	-.24***	-.09	-.10	-.21**	-.16*	-.25***	-.24***	-.08	.13	.30***
Total Group	-.09	-.05	-.15**	-.05	-.03	-.13**	-.11*	-.18**	-.14*	-.05	.05	.16**

* $p < .10$

** $p < .05$

*** $p < .01$

Table 7

Correlations Between Measures of Organization in Recall and Verbal-Ability for Males and Females

List Imagery and Sex of Subject	Intertrial Repetitions (ITR)			Sequential Consistency			Intertrial Correct Responses (ITC)			Intertrial New Responses (ITN)		
	1-2	2-3	3-4	1-2	2-3	3-4	1-2	2-3	3-4	1-2	2-3	3-4
<u>High Imagery List</u>												
Males	-.05	-.00	.11	.01	.02	.09	-.11	-.13	-.14	-.21	.18	.01
Females	-.03	-.01	-.11	.07	.00	-.09	.01	-.01	-.16*	.01	-.10	.08
Total Group	-.03	-.01	.02	.02	.00	.02	-.02	-.04	-.11*	-.10	.05	.02
<u>Medium Imagery List</u>												
Males	.05	-.00	-.08	-.00	.04	-.03	-.02	-.11	-.07	-.09	-.01	-.03
Females	-.11	-.08	-.21**	-.10	-.08	-.18**	.01	-.02	-.07	-.09	-.00	-.04
Total Group	.01	.01	-.04	.00	.07	.02	-.02	-.03	-.06	-.02	-.01	
<u>Low Imagery List</u>												
Males	.16	-.14	.08	.14	-.13	-.10	-.07	-.04	-.02	.11	.12	-.03
Females	-.04	.03	.03	-.05	.04	.05	-.24***	-.26***	-.29***	.01	.16	.28***
Total Group	.07	-.10	-.04	.06	-.08	-.01	-.10	-.08	-.07	.08	.14**	.09

*p <.10

**p < .05

***p < .01

significance ($p < .10$) only on the comparison of ITC data based on the last two trials. Thus, there is at least partial assurrance that inferences concerning the use of imagery strategies are more accurate when discussed in the context of imagery ability than of verbal ability.

Similar analyses conducted separately for each sex indicate that the relationships described above are almost exclusively limited to female Ss. Furthermore, in many cases correlations of similar magnitude were found for relationships between verbal-ability and organization as for relationships between imagery-ability and organization although there were half again the latter comparisons as there were for the former. The correlation between verbal-ability and imagery ability for sub-groups was .16.

Discussion

It is clear that variations in the quality of concreteness attributable to words in a list are related directly to the facilitation of performance and recall in learning by the study-recall procedure. This finding is, of course, a replication of that obtained in earlier studies such as those by Stewart (1965), Tulving, et al. (1965), and Paivio (1965). It was also replicated within the present study where two sets of tasks were employed, each representing slightly different norms. Accordingly, although the stimulus materials were presented in a single modality, the higher the imagery-provoking value or picturability of the materials to be learned the easier they are to recall. This effect is analogous to the finding that pictures are recalled more easily than verbal materials when concreteness is held constant as found by Stewart (1965). Thus, beyond the mere replication of the relation between concreteness and performance, the importance of these data is that they imply a process by which Ss employ pictorial representations, similar to that suggested by the term "ikonic-imagery," as a strategy for encoding certain types of material.

The main effect attributable to sex differences is impressive only because it emerges so consistently in studies where tasks of the sort employed in the present one are used. The finding that the performance of women Ss was superior to that of men Ss is undoubtedly related to the effects of differential cultural experiences on the acquisition of knowledge and strategies for learning and recall by men and women. Such differences are reflected in differential performance on tests of general intellectual ability (Broverman, Klaiber, Kobayashi, & Vogel, 1968).

In this study there was no evidence for learning to learn. Previous evidence in similar investigations has been controversial. Both Stewart (1965) and Tulving, et al. (1965) obtained increments in performance over lists which they reasoned were due to learning sets. Dallett (1963), too, obtained the effect in his studies. Murdock (1960), on the other hand, concluded from his independent investigations, that neither learning to learn or warm-up effects were to be observed in multi-trial free recall learning tasks. In reviewing their results, Tulving, et al. (1965) suggest that, perhaps, learning to learn occurs only where uncued free-recall, rather than cued recall as in Murdock's experiment, was employed. However, this explanation cannot be applied to the present experiment since cues were not used in the recall phase. Learning to learn would appear to be a theoretically reasonable outcome of the present procedure. Nevertheless, the findings from the present study together with the fact that Stewart (1965) obtained only very small differences in her study, suggest that the effect must be considered a fragile one at best under the conditions of this experiment.

There were two interactions between Imagery and treatments that were of importance in this study. First, low imagers perform better over all trials than do high imagers. This observation replicates a comparable one observed by Stewart (1965). However, in the present study, there was an ordinal interaction between imagers and trials. Accordingly, while low imagers perform significantly better than high imagers on the first two trials their performances do not differ significantly on the last two trials. Apparently, high imagers employ the early stages of learning for identifying means by which their preferred strategy of imagery can be implemented in that task. A parallel finding by Paivio (1969; Ernest & Paivio, 1969; Ernest & Paivio, 1971) indicates that individual differences in imagery were predictive of incidental memory, indicating differences in learning strategies of high and low imagers. Thus, he found that high imagers recalled more incidental components of a compound stimulus or response item than did low imagers but the two groups did not differ in intentional learning. The evidence accumulating to date suggests that low imagers and high imagers employ quite different strategies in studying and recall with the consequent effect on rate of learning and on the amount and nature of what is recalled.

The second interaction of importance was that between Imagers and levels of concreteness represented by words in a given list. A series of t-tests among means indicated that high imagers had difficulty in learning a list of low concreteness (i.e., of abstract words). However, their performance increased significantly with increases in concreteness. Again, this finding replicates one obtained by Stewart (1965). Furthermore, Ernest & Paivio (1969) also reports similar sex differences, finding better incidental memory for high imagery females but not for males. As Paivio (1970) suggests, "It is too early to say what this might mean, but developmental factors certainly must be implicated."

Of interest in the study of aptitude by treatment interactions is the finding that organization is not manifest until the last trials on the more difficult (abstract) list and is manifested on the first trials of the easier (concrete) list. However, the implications of each differ: they suggest that there are more or less capricious attempts at organization on the first trials of the abstract list by both groups, while both groups (i.e., high and low imagers) organize the concrete list about equally after the first two trials. The correlations based on ITC data demonstrate differences in the strategies employed by the two groups somewhat more definitively. Thus, high imagers achieve significantly ($\underline{p} < .10$) less organization than do low imagers on all comparisons of adjacent trials of the lists comprised of words with low-rated imagery, and on the comparison of the adjacent trials of the first three trials of the lists comprised of medium rated imagery. The degree of imagery-ability is unrelated to performance on any of the trial comparisons for the lists comprised of words rated high on imagery.

These data suggest that the high imagers may attempt to employ an inappropriate strategy for the abstract words thus hindering their performance on the low rated-imagery lists of words and to a lesser extent on the medium imagery words. On the other hand, the imagery strategy may be as appropriate for organizing the high imagery list as is any other strategy employed by Ss with low imagery ability. Or, perhaps, both groups employ the same strategy when learning lists of concrete words. In either case, the correlations are not significantly different from zero for the concrete list. In general, the correlation data lead to the conclusion that while imagery ability is not the exclusive factor involved in organization in recall, it is no less important than verbal ability.

These data reported here, together with those from other studies, further suggest that the preferred method of high imagers for encoding materials (that is, imaging) is inappropriate for low-concreteness words thus impairing their performance. However, the method is an efficient one when applied to materials that represent concrete referents or that can be imagined (pictured) easily. Accordingly, their (i.e., high imagers) performance improved in correspondence with increases in concreteness. The low imagers, who are hypothesized to employ other learning strategies, were relatively unaffected by changed in concreteness.

References

Bennett, G. K., Seashore, H. G., & Wesman, A. G. Differential aptitude tests. (Grades 8-13 and adults). New York: Psychological Corporation, 1963.

Bousfield, A. K., & Bousfield, W. A. Measurement of clustering and of sequential constancies in repeated free recall. Psychological Reports, 1966, 19, 935-942.

Broverman, D. M., Klaiber, E. L., Kobayashi, Y., & Vogel, W. Roles of activation and inhibition in sex differences in cognitive abilities. Psychological Review, 1968, 75, 23-50.

Cronbach, L. J., & Snow, R. E. Individual differences in learning ability as a function of instructional variables. Final Report, U. S. Office of Education, Contract No. OEC 4-6-061269-1217. Stanford, California: School of Education, Stanford University, 1969.

Dallett, K. M. Practice effects in free and ordered recall. Journal of Experimental Psychology, 1963, 66, 65-71.

Di Vesta, F. J., Ingersoll, G., & Sunshine, P. M. A factor analysis of imagery tests. Journal of Verbal Learning and Verbal Behavior, 1971, in press.

Ernest, C. H., & Paivio, A. Imagery and sex differences in incidental recall. (Research Bulletin No. 121) London, Ontario: Department of Psychology, University of Western Ontario, 1969.

Ernest, C. H., & Paivio, A. Imagery and verbal associative latencies as a function of imagery ability. Canadian Journal of Psychology, 1971, 25, 83-90.

Fagan, J. F. Measuring recall II: The ITR score expressed as a ratio. Psychonomic Science, 1968, 11, 205.

Mednick, S. A., & Mednick, M. T. Remote associates test. Boston: Houghton-Mifflin, 1967.

Murdock, B. B., Jr. The immediate retention of unrelated words. Journal of Experimental Psychology, 1960, 60, 222-234.

Noble, C. E. An analysis of meaning. Psychological Review, 1952, 59, 421-230.

Paivio, A. Abstractness, imagery, and meaningfulness in paired-associate learning. Journal of Verbal Learning and Verbal Behavior, 1965, 4, 32-38.

Paivio, A. Mental imagery in associative learning and memory. Psychological Review, 1969, 76, 241-263.

Paivio, A., Yuille, J. C., & Madigan, S. A. Concreteness, imagery, and meaningfulness values for 925 nouns. Journal of Experimental Psychology Monograph Supplement, 1968, 76, No. 1, Part 2.

Thorndike, E. L., & Lorge, I. The teacher's word book of 30000 words. New York: Bureau of Publications, Teachers College, Columbia University, 1944.

Thurstone, L. L., & Jeffrey, T. G. Space thinking (Flags). Chicago: Education-Industry Service, 1959.

Tulving, E., McNulty, J. A., & Ozier, M. Vividness of words and learning to learn in free recall learning. Canadian Journal of Psychology, 1965, 19, 242-252.

A Factor Analysis of Imagery Tests

Francis J. Di Vesta, Gary Ingersoll and Phyllis Sunshine

The mentalistic sounding construct "imagery" has gained renewed acceptance on several experimental fronts during the past decade and undeniably has led to a number of fruitful insights on learning processes. Despite its many definitions (Holt, 1964), recent investigators have found that to provide this concept with operational meaning, either by reference to the _S_'s behavior or by inference from objective descriptions of experimental conditions, was a relatively straightforward matter.

Several orientations in studies of imagery can be identified in the current literature. In one, represented by the work of Paivio (for example, see Paivio, 1970), the concern has been with the effects of stimulus characteristics, such as the perceived vividness or concreteness of events, usually language symbols, on the efficiency of learning processes. In another, the emphasis is on imagery as a _process_ or _strategy_ by which experiences become encoded or transformed for storage in, or later retrieval from long-term memory. Bower's (for example, see Bower, 1970) research is an illustration of this category of investigations. Within a third framework, imagery has been defined in terms of individual differences based on the subjects' reports of vividness of imaginal experiences (for example, Galton, 1880, 1883; Richardson, 1969) or on scores for objective tests (for examples, see Thurstone, 1944).

A potentially fruitful extension of the aforementioned orientations may exist within the aptitude by treatment interaction (ATI) framework proposed by Cronbach (1957) as a general methodology for blending the experimental and correlational methods. It was this orientation that provided the impetus for the present study. Examples of the application of ATI to the study of imagery are to be found in investigations described by Hollenberg (1970) and Stewart (1965). Only a few such studies have been conducted to the present. Nevertheless, it is becoming increasingly apparent (Paivio, 1970; Rohwer, 1970) that hypotheses related to the differential effects on performance of manipulated variables as they interact with differences in ability (or preference) to use ikonic imagery versus verbalization strategies in thinking are gaining attention.

An essential requirement for studies within the ATI orientation is a reliable and valid measure of imagery. Introspection was the basis for the earlier measures proposed by Galton and was retained in the more recent summary by Richardson (1969) who described the revised scales originally employed by Betts (1909) and Gordon (1949). A skeptical view of the reliability of self-reports of imagery and the consequent search for more objective measures prompted Hollenberg, in her investigations of visual imagery with children, and Stewart, in her studies with college students, to reject the self-report procedure and to employ spatial manipulations tests instead.

Both Hollenberg (1970) and Stewart (1965) made some further assumptions that influenced not only the choice of their tests but also the interpretation of the scores. They reasoned that individual differences in thinking by the use of images (visualizers), on the one hand, or by the use of language symbols (verbalizers), on the other, were the products of the individual's unique history of rewards and punishments for employing a given strategy. Permissiveness in child-training was believed to influence the continued use of imagery and of language habits related to imagery. Guidance in the use of symbols not directly based on perceptual similarities was reasoned to encourage the use of verbalizations in thinking. As the child begins to think through the use of symbols, ikonic imagery was assumed to fade; that is, verbalization is substituted for imagery. These assumptions led both investigators (Hollenberg, 1970; Stewart, 1965) to employ the spatial manipulations tests as though imagery was inversely related to verbal ability; that is, they assumed high scores represented imagers and low scores represented verbalizers. This inference was supported, in part, by the disordinal interactions found between treatments and aptitudes. Thus, high-imagery subjects (visualizers) tended to perform more effectively than low-imagery subjects on tasks hypothesized to favor ikonic mediation while low-imagery subjects (verbalizers) outperformed the high imagery subjects on tasks that favored verbal mediation.

This brief review suggests the hypotheses that (a) measures of imagery and verbal ability are independent (orthogonal), if not bi-polar, factors; and (b) measures of imagery based on introspection are different from and/or less reliable than measures based on objective tests. These hypotheses, of course, are directed toward an examination of the construct validity (Cronbach & Meehl, 1955) of imagery as an individual difference variable.

One means of testing these hypotheses is to determine the relationships between the three kinds (introspective reports of imagery, objectively defined tests of spatial manipulation, and tests of verbal ability) of tests.

to the performance of Ss on tasks assumed to be facilitated differentially by verbal and imaginal processes. This approach was the one employed by Hollenberg and Stewart. However, attempts to replicate Stewart's experiments in our laboratory were only partially successful. Accordingly, following Barratt (1953), the alternative procedure of testing these hypotheses via factor analytic procedures was used, and is the subject of this report. The underlying simple structure represented in several tests was examined in two separate studies. In the first study, the hypothesis related to the independence of verbal ability and imagery was examined; in the second study the relationship between introspective and objective tests of imagery was investigated.

Method

Subjects

There were 184 Ss in Study I and 232 Ss in Study II. All Ss were enrolled in an introductory course in educational psychology. Though participation was voluntary, Ss received credit toward their course grade for participating in the study. Most Ss had taken part in one or more experiments prior to enrollment in this course.

Materials

Study I. The battery of tests for Study I was purposely contrived to consist of at least two factors: One group of tests was hypothesized to depend primarily on perceptual skills or spatial manipulation. The second group consisted of tests related to verbal and general intellectual abilities. Another group, comprised of general personality variables, was included to prevent restrictions on the extraction of factors. The specific tests in this battery are described immediately below.

The Gottschaldt Figures Test as described by Thurstone (1944) consisted of 61 test items divided into five parts. There were 27 items in Part A, 7 items in Part B; 7 items in Part C; 10 items in Part D; and 10 items in Part E. Maximum time limits were two, one, three, four, and four minutes for each part, respectively. The score was the number of designs correctly traced within the time allowed.

The <u>Space Relations</u> test of the <u>Differential Aptitude Test Battery</u> (Bennett, Seashore, & Wesman, 1963) was employed as a measure of the ability to visualize the rotation of a picture or pattern in three dimensional space. The score for the total number of correct choices was obtained.

The <u>Space Thinking (Flags)</u> (Thurstone & Jeffrey, 1959) test was intended to measure an ability similar to that described for the <u>Space Relations</u> test. A time limit of 20 minutes was imposed. The score recorded was the number right minus the number wrong.

The <u>Stroop Color-Word Interference Test</u> (Stroop, 1935) was constructed according to standard procedures (for example, see Jensen and Rohwer, 1966; Thurstone, 1944). Three forms were administered to each <u>S</u>. In one the <u>S</u> read the names of colors printed in black. In a second, he named the colors of patches of color. The third version was the color-word interference task in which each word was printed in a color other than its color name. The total time for reading each version correctly was recorded.

The <u>Automatization Test</u> (Broverman, 1964; Broverman, Klaiber, Kobayashi, & Vogel, 1968) measured the <u>S</u>'s rate of naming three objects (tree, fly, and cup) repeated equally often on a card. There were depicted 110 representations of these objects on the card. The <u>S</u>'s score was the amount of time taken to name all objects correctly.

A <u>Vocabulary Test</u> was specifically devised for this study by modifying in multiple-choice form, several items selected from the <u>Henmon-Nelson Test of Mental Ability</u> (Lamke, Nelson, & Kelson, 1931-1960). This test consisted of nine verbal analogies, 17 vocabulary definitions or meanings, and four opposites.

The Scholastic Aptitude Test (SAT) of the College Entrance Examination Board (1962-1963) had been administered to all Ss prior to entrance to the university. Their verbal and mathematics scores on this test were obtained from their college records.

The Reading Comprehension test (Lindsay, Williams, & Peterman, 1969) was developed by the Student Affairs Research Office at the Pennsylvania State University. It consisted of 12 paragraphs, each of which was followed by two or more test items for each paragraph, for a total of 30 items. This test had been administered to the Ss during their Freshman year. Scores were obtained from their college records.

The Remote Associates Test (Mednick & Mednick, 1967) was administered as a test of ability to make mediating links in groups of words and, hence, was considered as a potential contributor to the verbal factor. It was administered with a 30-min. time limit. The score was the number of items answered correctly according to the key provided in the manual.

The personality measures consisted of the Achievement Anxiety Test (Alpert & Haber, 1960) from which debilitating anxiety and facilitating anxiety scores were obtained; the Tolerance for Ambiguity test devised by Budner (1963); and the Dogmatism Scale (Rokeach, 1960). These tests were devised, administered, and scored following descriptions provided in each of the references shown.

Study II. The test battery for Study II included the Space Thinking (Flags) Test; the Scholastic Aptitude Test (SAT); the Spatial Relations Test; and the Gottschaldt Hidden Figures Test, all of which have been described under Study I. In addition, the following were administered: the Digit Span test from the Wechsler Adults Intelligence Scale (Wechsler, 1955); the Tolerance for Ambiguity Scale (Budner, 1963); the Social Desirability Scale (Crowne, & Marlowe, 1964); the Memory-for-Designs Test (Graham & Kendall, 1960); the Betts Vividness of Imagery Scale (Betts, 1909); and the Gordon Test of Visual Imagery Control (Gordon, 1949). The latter two tests were employed as described in Richardson (1969). These tests, too, were devised, administered, and scored as described in the references cited.

Procedures

The tests for both studies were administered according to standardized instructions and procedures provided in manuals and references for the tests. There were, however, two exceptions: The items for the Digit Span Test were recorded and the items for the Memory-for-Designs tests were placed on 35-mm. slides, so that they could be presented to small groups rather than individually. The Stroop Color-Word Tests and the Automatization Test were administered individually to the S usually prior to or following participation in another experiment. All other tests were administered to Ss in small groups of 15-25 Ss which were monitored by two Es.

Results

In both Study I and Study II the basic data were the raw scores from the tests employed in each study. Pearson product-moment correlations among all scores within a study were calculated and then used in the principal components analysis for initial factorization. Six factors were extracted with eigenvalues greater than 1.00 in both studies. The factors extracted were rotated via the Varimax routine (Kaiser, 1958). Stability of the factor structure was achieved for the rotation of four factors in Study I and for three factors in Study II. The results of these analyses are described below.

Study I. The first study was concerned with the question of whether imagery and verbal ability were separable individual difference characteristics or whether they were constituents of a more general intellectual ability. The means and standard deviations of scores from the 20 aptitude tests are summarized in Table 1 separately for males and females. The intercorrelations among all variables for all Ss (N = 184) are displayed in Table 2. Note that sex was included as a "score" ("one" was employed for females and "two" for males) in a manner similar to that employed in Thurstone's (1944) earlier studies of perception. The authors were aware of difficulties associated with this practice but in view of the similarity in results obtained via separate analyses for each sex, it was decided that the most parsimonius means of presenting descriptive data for the entire group was by the summary of the analysis based on all Ss.

The results of the analyses are shown in Table 3. The first factor is comprised of Reading Comprehension, the Scholastic Aptitude Test, the Vocabulary Test, and the Remote Associates Test. These measures appear to be clearly classifiable as Verbal or Symbolic Imagery.

The second factor is comprised of the Flags, Spatial Relations, Gottschaldt, and the SAT:Math tests scores. This factor, with the exception of the SAT:Math test (which has the lowest of the loadings) can be defined as Ikonic Imagery.

The third component extracted was named Automatization. It was comprised of the three forms of the Stroop Color-Word Test and the Automatization test. Although it had been assumed that these tests might have had large "imaginal" components it is clear that the factor is separately defined from Ikonic Imagery thereby supporting results

Table 1

Means and Standard Deviations for Females and Males

On Twenty Measures: Study I

Test	Females (N = 104)		Males (N = 80)	
	$\overline{X}$	SD	$\overline{X}$	SD
Flags	106.40	19.98	115.50	12.84
Spatial Relations	68.39	16.77	76.35	15.91
Gottschaldt Figures Test: I	16.59	7.43	18.41	7.82
Gottschaldt Figures Test: II	5.28	2.02	5.80	1.85
Gottschaldt Figures Test: III	4.62	1.98	5.19	1.77
Gottschaldt Figures Test: IV	4.59	2.64	5.85	2.94
Gottschaldt Figures Test: V	6.39	2.04	7.30	2.05
Stroop: Word Score (secs.)	38.02	4.94	40.41	4.75
Stroop: Color Score (secs.)	52.29	7.03	56.83	7.84
Stroop: Interference Score (secs.)	93.21	15.69	100.30	19.04
Automatization	53.06	7.81	54.40	6.70
SAT: Math	549.50	83.07	569.00	78.41
SAT: Verbal	527.30	91.24	500.80	78.81
Vocabulary	19.13	3.73	18.23	3.97
Remote Associates Test	13.40	5.06	13.59	4.51
Reading Comprehension	16.21	4.35	16.24	4.62
Debilitating Anxiety	27.59	6.48	25.29	4.92
Facilitating Anxiety	24.36	4.80	26.01	4.07
Tolerance of Ambiguity	41.35	8.83	42.48	8.76
Dogmatism	128.90	19.14	130.70	18.00

Table 2

Matrix of Intercorrelations Among Variables in Study I

	Variables																		
Variables	SR	G:I	G:II	G:III	G:IV	G:V	S:W	S:C	S:CW	A	SATM	SATV	V	RAT	RC	DA	FA	TA	D	S

Flags	35	33	19	32	21	29	13	05	00	-01	24	-12	-10	06	10	-03	10	10	14	26
Spatial Relations		22	31	17	25	40	23	12	09	06	36	-04	04	08	02	02	-04	16	11	24
Gottschaldt: I			40	21	28	39	-06	04	-06	-13	24	11	05	-02	01	-11	11	10	08	12
Gottschaldt: II				41	38	47	02	00	-12	-10	38	22	06	12	-01	-07	07	07	00	13
Gottschaldt: III					55	44	09	-01	-08	-04	20	-01	-06	10	-05	-04	06	00	00	15
Gottschaldt: IV						59	04	01	-07	-03	29	10	09	15	01	-09	15	12	04	22
Gottschaldt: V							08	05	-04	00	37	13	07	14	08	-04	12	11	09	22
Stroop: Word								54	54	45	00	-29	-30	-05	-15	09	-11	04	11	24
Stroop: Color									71	60	05	-13	-16	-03	-11	04	-04	-01	11	29
Stroop:Color Word										59	-07	-21	-19	00	-13	04	00	-13	09	20
Automatization											-06	-12	-17	00	-15	10	-09	-12	05	09
SAT: Math												35	26	18	29	-14	17	03	01	12
SAT: Verbal													62	29	51	-24	20	-02	-16	-15
Vocabulary														33	48	-19	21	-11	-13	-12
RAT															11	-12	07	-03	-12	-02
Reading																-19	28	02	00	00
Deb. Anxiety																	-61	02	23	-19
Fac. Anxiety																		04	-14	18
Tol. of Ambiguity																			23	06
Dogmatism																				05
Sex																				

Table 3

Summary of Factor Loadings for Tests Associated with

Four Varimax Factors: Study I*

Test	I Verbal	II Imagery	III Automatization	IV Anxiety	h^2
Flags	-.12	.57	.08	.09	.42
Spatial Relations	.09	.58	.22	-.15	.46
Gottschaldt Figures Test: I	.03	.57	-.08	.08	.35
Gottschaldt Figures Test: II	.18	.67	-.06	-.04	.54
Gottschaldt Figures Test: III	-.06	.65	-.04	.03	.60
Gottschaldt Figures Test: IV	.10	.70	-.02	.07	.57
Gottschaldt Figures Test: V	.16	.77	.04	-.02	.63
Stroop Test: Word	-.21	.14	.73	-.08	.60
Stroop Test: Color	-.02	.04	.86	.00	.75
Stroop Test: Word/Color	-.07	-.11	.86	.03	.76
Automatization	.00	-.11	.78	-.10	.65
SAT: Math	.50	.49	.06	.04	.51
SAT: Verbal	.83	.02	-.17	.11	.73
Vocabulary	.81	-.05	-.17	.10	.70
Remote Associates Test	.50	.10	.06	.02	.32
Reading Comprehension	.63	-.00	-.10	.20	.61
Debilitating Anxiety	-.15	-.05	.04	-.85	.74
Facilitating Anxiety	.15	.11	-.03	.83	.74
Tolerance of Ambiguity	-.15	.26	-.12	-.06	.42
Dogmatism	-.15	.19	.10	-.35	.51
Sex**	-.19	.37	.34	.39	.46
Eigenvalues	2.12	3.78	3.35	1.49	

* N = 80 males and 104 females.

** 1 = Females; 2 = Males.

obtained by Broverman (1964). This factor refers to the S's ability to perform simple repetitive tasks without being distracted by interfering influences, such as the general effect produced when the actual colors of the printing interfere with reading color-names.

The fourth factor obtained was Anxiety. As one might expect, this factor was comprised of the two scores, representing facilitating and debilitating anxiety, from the Achievement Anxiety Scale. This factor is interesting only because the two scores are differentially polarized on the factor, thereby providing a degree of validity for the constructs as hypothesized in the development of the scales (Alpert & Haber, 1960).

None of the loadings of the sex variable are high for any of the factors. Nevertheless, it is of interest to note the directions of the correlations: the Verbal factor was inversely correlated and the other factors were positively correlated with sex indicating that females tended to achieve higher scores on the Verbal factor than did males and males tended to achieve higher scores on the Automatization and Ikonic Imagery factors than did females. A similar tendency was noted by Thurstone (1944).

Study II. The analysis of data obtained for the second study was directed toward examining the validity of the notion that introspective reports and objective tests of imagery ability provide meas... of the same individual difference characteristics. The means and standard deviations of scores for all tests in this study are summarized in Table 4.

These data are based on the entire pool of Ss for Study II. A comparison with Study I indicates the results from the two studies were within a standard deviation of one another for overlapping tests,

Table 4

Means and Standard Deviations for Scores

On Ten Tests: Study II*

Test	$\overline{X}$	SD
Social Desirability	14.65	5.06
Vividness of Imagery	196.94	29.68
Control of Imagery	29.86	6.79
Digit Span	9.26	2.05
Memory for Figures	16.47	3.71
SAT: Verbal	509.47	88.04
SAT: Math	552.13	86.86
Gottschaldt Figures Test	36.10	11.89
Flags	108.93	19.98
Spatial Relations	67.72	17.68

* N = 232

although in each case the mean scores for Ss in Study I were higher than those for Ss in Study II. Thus, for Studies I and II, respectively, the SAT:Verbal scores were 515.80 and 509.47, the SAT:Math scores were 558.00 and 552.13, the Flags Test scores were 110.40 and 108.93, the Gottschaldt Figures Test scores wre 39.65 and 36.10, and the Spatial Relations Test scores were 71.85 and 67.72.

The intercorrelations among all variables in Study II are presented in Table 5. The correlations between the same variables in Study I and Study II were comparable except for the correlations between SAT:Verbal and the Spatial Relations test which were -.04 in Study I and .21 in Study II; and between SAT:Math and SAT:Verbal which were .35 and .54 in Study I and Study II, respectively.

The summary of the rotated factor matrix for Study II is shown in Table 6. The results presented there provide a clear reproduction of the Verbal and Imagery factors extracted in the factor analysis for Study I. As in Study I, women were found to be more facile in verbal than were the men and men were higher in imagery ability than were women. In addition, there was extracted a third factor described by the label Social Desirability. The constituents of this factor were the Control of Imagery Scale, the Vividness of Imagery Scale, and the Social Desirability Scale.

In summary, Study II provided a replication of the distinction between verbal and imagery abilities and, in addition, indicated that objective tests provide measures of abilities that may be quite different from those provided by introspective reports.

Table 5

Matrix of Intercorrelations Among Variables in Study II*

Variables	SD	VI	CI	DS	MFF	SATV	SATM	GFT	F	SR
Sex**	04	01	04	-11	00	12	-05	03	-22	-14
Social Desirability		29	20	-07	05	-12	-06	02	-03	-04
Vividness of Imagery			45	00	13	-06	00	06	-04	-04
Control of Imagery				03	07	09	14	09	-01	-08
Digit Span					02	22	18	02	05	07
Memory for Figures						15	28	26	24	38
SAT: Verbal							54	30	-06	21
SAT: Math								44	25	29
Gottschaldt Figures Test									28	34
Flags										37
Spatial Relations										

*Decimal points have been omitted in correlation coefficients.

**In this study men were assigned a score of one and women were assigned a score of two.

Table 6

Summary of Factor Loadings for Tests

Associated with Three Varimax Factors: Study II

Tests	Factor I Verbal	Factor II Imagery	Factor III Social Desirability	h^2
Flags	-.02	.78	.04	.61
Spatial Relations	.26	.70	.08	.56
Gottschaldt Figures Test	.51	.44	-.16	.48
Memory for Figures	.26	.53	-.21	.40
Control of Imagery	.17	-.08	-.75	.59
Vividness of Imagery	-.04	.01	-.81	.66
Social Desirability	-.20	.03	-.62	.43
SAT: Math	.75	.32	-.06	.67
SAT: Verbal	.86	-.07	.04	.76
Digit Span	.37	-.01	.09	.14
Sex*	.26	-.47	-.12	.31
Eigenvalue	1.40	2.52	1.70	

* In this study men were assigned a score of one and women were assigned a score of two.

107

Discussion

The results of this study clearly confirm the stability of imagery as an individual difference variable. In large part, the tests comprising this factor have no apparent dependence on obvious meaningful associations. An analysis of manipulations required for each set of tasks would imply that they could only be accomplished with minimal benefit from verbal associations, labeling, or conceptualizations. Furthermore, it appears that successful performance on these tasks required those mental abilities with imaginal rather than symbolic properties.

The processes, presumably indexed by the imagery factor can be inferred from an analysis of the kinds of skills required to perform the tests which saturate this factor in both studies. The ability to hold a percept in memory long enough to work with it was measured by the Memory-for-Designs test. The Flags and Spatial Relations tests require the S not only to hold the percept in memory but to rotate it or to unfold it in various ways, that is, to "view" it from many perspectives. However, these are relatively primitive abilities. The measure of the higher forms of imagery would require a task in which the figure would be changed or reorganized. To some extent this function was served by the Gottschaldt Figures Test. In these tasks, the critical figure to be identified was camouflaged by extraneous lines. In order to perceive the hidden, less obvious stimulus relationship the initial more obvious percepts had to be subdued. Thus, the underlying process in the Gottschaldt test appears to be one of not only holding the stimulus in memory but also of restructuring the percept.

The present factor analysis, of course, only implies that the Flags, Spatial Relations, Memory-for-Designs and Gottschaldt tests comprise a cluster of tasks that presumably require some common process. It does not, except by inference, indicate whether imaginal, visualization, perceptual, or some other process underlies performance of the tasks involved. An integration of a finding from Barratt's (1953) study, with the findings from the present study, may help to provide a partial answer to this question. He required his Ss to perform sample tasks representing each factor extracted and then to rate their use of imagery during the performance of these tasks. It was found that high-imagers performed better than low-imagers on the Spatial Manipulations tasks but the performance of the two groups on the Spatial Recognition and Spatial Reasoning tasks was not differentiated. Barratt indicates that these results justify the use of these tests for measuring imagery. In view of the cross-validation achieved by differences in performance of the two imagery groups on the tasks representing the separate factors his reasoning is warranted, at least in part. Nevertheless, in view of the present findings regarding the possibility that introspective reports are partially confounded with social desirability, complete answers to the validity of the imagery construct can only come from further development of a carefully constructed nomological net, a part of which must necessarily be comprised of the findings from Barratt's and from Stewart's studies.

The separate extraction of the Automatization factor from the imagery factor, was in a sense, predictable from Broverman's (1964) studies. On the surface it may appear that both the Gottschaldt Figures Test and the Stroop Color-Word Test are similar to the extent that they involve interference of performance by the presence of irrelevant stimuli. However, the distinctiveness of the tasks lies in the relationship between the interfering stimuli to the critical stimuli in each. Accordingly, in the Stroop Color-Word Test the stimulus attribute is readily perceived. Successful performance is dependent on a set to respond to certain obvious features of the stimulus and not to respond to the wrong, interfering, though equally obvious features. This <u>set-to-respond</u> in certain ways is undoubtedly present in Automatization. It is clearly distinct from the reorganization of stimulus structure required in Imagery, as represented in the Gottschaldt test.

The Verbal factor identified in both studies is so familiar that very little additional explanation to that already provided in textbooks and manuals seems necessary. As one can readily see, it is the factor comprised of acquired verbalizations and language symbols and the ability to employ these symbols, in various ways, within tasks where general mental ability and symbolic transformations of incoming stimuli facilitate performance. Similarly, the extraction of the Anxiety factor in Study I was not unanticipated since, in two ways, it was quite unlike any of the other measures. First, it was a self-report or introspective measure rather than an objective test and second, the questions related to the affective domain of behaviors rather than to the cognitive domain. Nevertheless, the extraction of the Verbal and Anxiety factors are important to the present discussion to the extent that, except for the loading of the SAT:Math score on Imagery, none of the other loadings on these factors overlapped significantly with those of Imagery. Thus, the data provide further evidence that Imagery is a separate constituent of cognitive structure or a separate cognitive strategy from Verbal ability.

The final factor to be discussed is Social Desirability. It was not altogether surprising that one component of this factor, the introspective reports of ability to control imagery and vividness of imagery loaded heavily on a factor other than that represented by the objective tests of imagery. However, it was surprising that they should be heavily weighted on Social Desirability. Inspection of the items on these scales suggest a possible explanation: The scales for the introspective reports of imagery may imply to some Ss that to be able "to control one's imagery" and "to experience vivid images" is a highly valued characteristic comparable, for example, to possessing a high I.Q. The Social Desirability scale also measures qualities of behavior that reflect dependence on the acceptance, recognition and approval of others. If ability to conjure up images is believed to be a culturally desirable trait then it is consistent that this bias will affect the scores on the scale. It is most interesting to note that Richardson (1969, p. 87) does indicate a correlation between the richness of fantasy (measured by introspective reports) and persuasibility (defined as readiness to accept social influence regardless of what is known about the communicator). However, nowhere in the book could the present investigator find where response bias, need for approval, or similar behavioral qualities were considered in the interpretation of data presented. Nevertheless, it was clear that differential performances such as "perceptual achievements, ... involved in responding to an ambiguous ink blot or in recognition of an object" (Richardson, 1969, p. 131) and other similar behaviors attributed by Richardson to differences between high and low visualizers could also be influenced by social desirability and thereby would provide alternative explanations.

In summary, the results of the present investigation appear to warrant two conclusions. First, imagery as defined by objective tests appears to be a distinctive individual difference variable. Relating underlying processes such as those tentatively described above to performance in situations predominantly concrete or abstract should be a fruitful source of hypotheses for further investigations. In this context, the present investigator wishes to reemphasize the caution certainly familiar to potential investigators, that is, hypotheses should be based on processes assumed to underlie the factors rather than on the labels attached to the factors. At best, such labels reflect the idiosyncracies of the investigator and by themselves are more often misleading than not. A case in point is the ambiguity associated with a term such as automatization which can be interpreted variously as "readiness to respond," "rigidity," "set to respond to given attributes," or even perhaps as "fluency in translating pictures into words." We have already pointed out in the introduction the numerous definitions attached to imagery. The term Imagery, even as it is used here, does no more than imply a non-linguistic category. The question of whether it is an ikon, engram, or non-linguistic meaning category is certainly unresolved. Our present inclination is to restrict its definition to those processes presumably involved in test performance. Still further refinement of these definitions appear to be imperative if the elusive aptitude by treatment interactions are to be captured in systematic investigations.

A second conclusion from the present study is that introspective reports, as measures of imagery, do not possess construct validity. This conclusion was implied not only by the results of the present investigation but by the examination of data presented in reports of studies where introspective scales were employed. Where such introspective scales have been employed, as they were in studies reported by Richardson (1969), the data crediting differences in performance to differences in imagery should be interpreted with caution. Some consideration may be given to further development of these scales with an attempt to remove response bias or their heavy dependence on social desirability. For the present, they must be considered to be confounded with response bias.

This study provides only a description of one structure of imagery. It does not answer important and interesting developmental questions such as the degree of imagery relative to verbal ability in children compared to adults. Nor does it indicate how much the use of imagery reflects a skill as it probably does with adults. These kinds of distinctions suggest interaction with manipulations of task characteristics (such as concreteness) in the former and with manipulations of motivational levels in the latter. The interpretations provided here should provide a basis for further factorial studies to differentiate among imagery structures or specific kinds of imagery that may vary for the senses. However, the primary interest is to determine whether imagery as isolated here is descriptive of intellectual performance that transcends purely perceptual effects. In view of the potential fruitfulness of the approach and the current popularity of imagery as a cognitive process, investigations of the antecedent conditions associated with it and of its interactions with task and stimulus variables are in order.

References

Alpert, R., & Haber, R. N. Anxiety in academic achievement situations. Journal of Abnormal and Social Psychology, 1960, 61, 207-215.

Barratt, P. E. Imagery and thinking. Australian Journal of Psychology, 1953, 5, 164-174.

Bennett, G. K., Seashore, H. G., & Wesman, A. G. Differential aptitude tests. (Grades 8-13 and adults.) New York: Psychological Corporation, 1963.

Betts, G. H. The distributions and functions of mental imagery. New York: Teachers College, Columbia University, 1909.

Bower, G. H. Analysis of a mnemonic device. American Scientist, 1970, 58, 496-510.

Broverman, D. M. Generality and behavioral correlates of cognitive styles. Journal of Consulting Psychology, 1964, 28, 487-500.

Broverman, D. M., Klaiber, E. L., Kobayashi, Y., & Vogel, W. Roles of activation and inhibition in sex differences in cognitive abilities. Psychological Review, 1968, 75, 23-50.

Budner, S. Tolerance of ambiguity as a personality variable. Journal of Personality, 1963, 30, 29-50.

College Entrance Examination Board Scholastic Aptitude Test. Princeton, New Jersey: Educational Testing Service, 1962-1963.

Cronbach, L. J. The two disciplines of scientific psychology. American Psychologist, 1957, 12, 671-694.

Cronbach, L. J., & Meehl, P. E. Construct validity in psychological tests. Psychological Bulletin, 1955, 52, 281-302.

Crowne, D. P., & Marlowe, D. The approval-motive: Studies in evaluative dependence. New York: Wiley, 1964.

Galton, F. Statistics of mental imagery. Mind, 1880, 5, 301-318.

Galton, F. Inquiries into human faculty and its development. London: J.M. Dent & Sons, 1883. Reprinted New York: Everyman, 1911.

Gordon, R. An investigation into some of the factors that varies the formation of stereotyped images. British Journal of Psychology, 1949, 39, 156-167.

Graham, F. K., & Kendall, B. S. Memory-for-Designs Test: Revised General Manual. Perceptual and Motor Skills, 1960, 11, 147-188.

Hollenberg, C. K. Functions of visual imagery in the learning and concept formation of children. Child Development, 1970, 41, 1003-1015.

Holt, R. R. Imagery: the return of the ostracized. American Psychologist, 1964, 19, 254-264.

Jensen, A. R., & Rohwer, W. D., Jr. The Stroop color-word test: A review. Acta Psychologica, 1966, 25, 36-93.

Kaiser, H. F. The Varimax criterion for analytic rotation in factor analyses. Psychometrika, 1958, 23, 187-200.

Lamke, T., Nelson, M. J., & Kelson, P. C. The Henmon-Nelson tests of mental ability. Revised edition. Boston: Houghton-Mifflin, 1931-1960.

Lindsay, C. A., Williams, G. D., & Peterman, D. J. Technical manual for the 1969 Division of Counseling pre-registration testing program. University Park, Penna.: Pennsylvania State University, 1969. Mimeo.

Mednick, S. A., & Mednick, M. T. Remote associates test. Boston: Houghton-Mifflin, 1967.

Paivio, A. On the functional significance of imagery. Psychological Bulletin, 1970, 6, 385-392.

Richardson, A. Mental imagery. New York: Springer, 1969.

Rohwer, W. D., Jr. Images and pictures in children's learning: Research results and educational implications. Psychological Bulletin, 1970, 6, 393-403.

Rokeach, M. The open- and closed-mind. New York: Basic Books, 1960.

Stewart, J. C. An experimental investigation of imagery. Unpublished doctoral dissertation, University of Toronto, 1965.

Stroop, J. R. Studies of interference in serial verbal reactions. Journal of Experimental Psychology, 1935, 18, 646-660.

Thurstone, L. L. A factorial study of perception. Chicago: University of Chicago Press, 1944.

Thurstone, L. L., & Jeffrey, T. G. Space thinking (Flags). Chicago: Education-Industry Service, 1959.

Wechsler, D. Wechsler Adult Intelligence Scale. New York: Psychological Corporation, 1955.

Imagery Ability, Abstractness and Word Order

As Variables in Recall of Adjectives and Nouns

Francis J. Di Vesta and Steven M. Ross

Among the empirical results of experiments on item imagery-arousal are the findings that word pairs are more easily learned in the noun-adjective order as opposed to adjective noun order (Lambert & Paivio, 1956; Paivio, 1963; Kusyszyn & Paivio, 1966; Yuille, Paivio & Lambert, 1969); that concrete nouns, but not abstract nouns are more effective associative cues than adjectives (Lockhart, 1969); and that pictures are easier to remember than concrete nouns (Stewart, 1965; Paivio, 1969). Although investigations of imagery as a stimulus attribute has yielded generally consistent and reasonably conclusive evidence, there is considerably less empirical support for theoretical notions regarding imagery as a transformational process influenced by individual differences in imagery-ability.

A potentially fruitful approach in defining imagery abilities has involved the use of spatial manipulation tasks. Using these measures, Ernest and Paivio (1969; 1970) found that female high imagers are more accurate in recalling incidental material than female low imagers, and that high imagers of both sexes have greater reaction times in associating abstract stimuli than in associating concrete stimuli. Stewart (1965) too, has reported that the performance of high imagers was facilitated by concrete stimulus materials in several learning situations.

The present investigation was, in part, a replication of one by Yuille, Paivio, and Lambert (1969) in which the order of presenting paired-associates

117

was manipulated; that is, Ss were presented pairs of words in either the noun-adjective or adjective-noun order. In addition, controls were provided for the abstract-concrete characteristics of the adjectives as well as of the nouns. It was found that recall was superior for the noun-adjective order, for high-Imagery rather than low-Imagery stimulus elements, and for high-Imagery rather than low-Imagery response elements. Overall, stimulus Imagery emerges as the most critical factor in accounting for differences in item recall.

The main purpose of the present study, however, was to extend the one conducted by Yuille, et al., (1969) by incorporating levels of imagery, as an individual difference variable. This factor was included in order to investigate the hypothesis that levels of imagery ability interact with treatments to affect performance. Specifically, it was hypothesized that, relative to low imaging ability, high imagery ability would facilitate the learning of abstract stimuli more than of concrete stimuli. These hypotheses were suggested in several studies by Paivio and his colleagues. For example, Paivio and Foth (1970) demonstrated that mediation instructions emphasizing imaginal processes resulted in greater recall of abstract pairs than mediation instructions emphasizing verbal processes. Ernest and Paivio (1971) showed that the relative superiority of high imagers over low imagers in reaction speed was greater when the stimuli to be associated were abstract for both imaginal and verbal instructions. In general, these findings suggest that imaginal processes complement verbal associations with the presentation of abstract stimuli. Concrete stimuli, on the other hand, elicit easily detectable cues that are as accessible for dual processing along with verbal cues by high imagers and non-imagers alike. However, non-imagers are presumably less able to establish pictorial associations and, as a result, are less efficient at retrieving the particular response when abstract stimuli are used. Therefore it was hypothesized that high imagers would have relatively greater recall for abstract pairs than would low imagers.

Method

Design

The Ss were presented two study-recall trials of a list of paired-associates. The word pairs appearing in a given list were presented in either a noun-adjective (N-A) or adjective-noun (A-N) order. Thus, in one condition, nouns served as the stimulus elements and adjectives as the response elements. In the other condition, adjectives served as stimuli while nouns served as response elements. Imagery arousal (I) of the noun associates was employed as a within-subjects variable with two levels. Using the rating scales by Paivio, Yuille, and Madigan (1968), half of the nouns were selected on the basis of high imagery and concreteness, and the other half were selected on the basis of low imagery and abstractness. This variable was orthogonally crossed with a similar set of conditions in which the imagery arousal (I) of the adjective associates as another within-subjects variable was manipulated. Thus, within each order (N-A or A-N), a given list of word-pairs was comprised of equal numbers of word pairs in which both members of a pair had high rated imagery, in which one member had high rated imagery and the other had low rated imagery, and in which both members had low imagery. The between-subjects variable of list order was orthogonally crossed with another between-subjects variable, that of imagery ability as an individual difference variable. Thus, half of the Ss were classified as high imagers and the other half as low imagers according to their performance on spatial relations tests. All Ss were given two study-recall trials. In summary, these manipulations implied a 2 x 2 x 2 x 2 x 2 analysis of variance with the between-subjects factors being two orders of presentation (N-A and A-N), and two imagery aptitude groups (high-imagers and low-imagers); and the within-subjects factors being two levels of stimulus imagery (High-I and Low-I), two levels of response imagery (High-I and Low-I), and two study-recall trials.

Subjects

The S pool for this experiment consisted of approximately 300 undergraduates enrolled in an introductory educational psychology course at the Pennsylvania State University. Sixty-five high imagers and 65 low imagers were selected as potential Ss according to their average standard score (T) on a test battery of three spatial relations tests: Flags: A Test of Spatial Thinking (Thurstone & Jeffrey, 1956), A Space Relations Test from the Differential Aptitude Test (Bennett, Seashore, & Wesman, 1947), and the Gottschaldt Figures Test as described by Thurstone (1944). Of these, 54 high imagers (average T-score = 59.50) and 54 low imagers (average T-score = 37.75) agreed to participate in the present experiment, and were extended credit towards the course grade. The experimental variations were administered to Ss in groups of 3. All Ss were randomly assigned to separate conditions upon entry into the laboratory.

Word Lists

A stimulus list consisted of 24 adjective-noun (A-N) pairs; six pairs of which were comprised of high-imagery adjectives and high-imagery nouns (HH); six pairs were high-imagery adjectives and low-imagery nouns (HL); six pairs were low-imagery adjectives and high-imagery nouns (LH); and six pairs were low-imagery adjectives and low-imagery nouns (LL). Three randomized lists were prepared, thus requiring a pool of 36 adjectives and 36 nouns. Items for the pairs were randomly selected from this pool without replacement. The noun-adjective (N-A) lists were formed by reversing the pairs in the A-N lists. Altogether six lists were prepared, each with two random orders of presentation. No word-pair occurred more than once in either the A-N or N-A Presentation Order.

The 18 high-imagery (concrete) nouns had average ratings of 6.40 on the imagery scale and 6.80 on the concreteness scale; the 18 low imagery (abstract) nouns had imagery and concreteness ratings of 3.50 and 1.77 respectively. These words were selected from the Paivio, Yuille, and Madigan (1968) norms. Separate norms for a set of 75 adjectives and five nouns were obtained specifically for this study inasmuch as ratings of adjectives on the imagery and concreteness scale were not readily available. The five nouns were selected at random from the Paivio et al. (1968) norms in order to provide a basis for determining similarity in ratings for the two groups. The set of 75 adjectives and five nouns were rated for imagery arousal by 15 undergraduate volunteers who were from the same S pool as that employed for the experiment. On the basis of their ratings, 18 High-I adjectives averaging 6.29 on a 7-point scale, and 18 Low-I adjectives averaging 2.74 were selected. The average ratings of the five nouns was 4.22 which did not differ significantly from the 4.02 obtained by averaging the corresponding I scores from the Paivio et al. (1968) norms. Thus, it was assumed that the ratings of adjectives were comparable to the ratings of nouns available from Paivio, et al.'s norms.

Following the selection of the individual words, each adjective was paired with both a High-I and a Low-I noun. These pairs are displayed in Table 1. Each word-pair was printed on a 3 x 5 in. card.

Table 1

Nouns and Adjectives used in the Paired-Associate Lists

Adjectives*		Nouns	
High Imagery	Low Imagery	High Imagery	Low Imagery
Dark	Boring	Elephant	Satire
Hairy	Known	Fox	Spirit
Round	Bad	Camp	Edition
Bumpy	Subtle	Hurdle	Crisis
Vertical	Mature	Acrobat	Quality
Rocky	Usual	Volcano	Disposition
Fuzzy	Trite	Clothing	Jealousy
Blue	Hungry	Reptile	Obsession
Sharp	Best	Corner	Anger
Colorful	Actual	Factory	Honor
Cloudy	Personal	Alcohol	Belief
Glassy	Different	Skillet	Pleasure
Smooth	Popular	Jelly	Intellect
Shiny	Quiet	Revolver	Memory
Wet	Obvious	Barrel	Sensation
Burnt	Dominant	Headlight	Betrayal
Small	Tardy	Whale	Idea
Bloody	Real	Tweezers	Virtue

* Each adjective (High-I and Low-I) was paired with the corresponding High-I and Low-I nouns.

Procedure

Three Ss were scheduled for each experimental session and were seated at separate locations in the laboratory. Prior to Ss arrival at the session, one of the six decks of 3 x 5 inch cards had been placed randomly at each of the three positions. Thus, each S within a group received a unique deck consisting of two study-recall trials in either the A-N or N-A order.

All Ss were given standard-paired associate instructions administered by means of a tape recorder. The E was always present to supervise the general administration and to respond to questions regarding the procedure. The basic task was identical for all treatment conditions. At the sound of a bell, S viewed the first pair by flipping over the top index card. Ten seconds were allotted to study each pair after which S was again directed by the bell to proceed to the next card. Verbal instructions announced the end of the study list and the beginning of the recall session. During recall only the stimulus elements of the individual word pairs appeared on the index cards. An interval of 15 seconds was provided for S to write down the appropriate response element. Cover sheets were used to obscure all previous answers. When the recall trial was completed the answer sheet was collected and E proceeded to administer the second study-recall presentation of the same pairs via the same procedure employed for the first trial.

Results

The number of correct responses were analyzed by a mixed analysis of variance with two between-subjects and three within-subjects variables. The between-subjects factors were two Presentation Orders (N-A and A-N), and two levels of Imagery (high imagers and low imagers); the within-subjects factors were the two levels of Stimulus Imagery (High I and Low I) two levels of Response Imagery (High I and Low I), and two Recall Trials.

The significant main effects were as follows: The effect due to Presentation Orders yielded $\underline{F}$ (1,104) = 15.09, $\underline{p} < .001$, indicating that the N-A order ($\bar{X}$ = 4.36) was superior to the A-N order ($\bar{X}$ = 3.58) across conditions. The effect due to Stimulus Imagery yielded $\underline{F}$ (1,104) = 134.96, $\underline{p} < .001$. The effect due to Response Imagery yielded $\underline{F}$ (1,104) = 36.17, $\underline{p} < .001$. These findings imply that High-I words were more easily associated than Low-I words whether in stimulus or response positions. As would be expected the effect due to Recall Trials was highly significant yielding $\underline{F}$ (1,104) = 532.35, $\underline{p} < .001$. These main effects and the non-significant ($\underline{p} > .05$) effect associated with Imagery levels were qualified by the interactions discussed in the paragraphs that follow.

The effect due to the interaction between Presentation Order and Recall Trials yielded $\underline{F}$ (1,104) = 6.09, $\underline{p} < .05$. Though more words were correctly recalled during the second recall trial for both the N-A and A-N orders, the degree of improvement was more pronounced for the A-N condition. This effect is almost certainly attributable to the near ceiling performance of $\underline{S}$s in the N-A condition during the first recall trial ($\bar{X}$ = 4.68). If this is contrasted with the average recall in the A-N condition ($\bar{X}$ = 2.48), it is obvious that comparative potentialities for improvement were markedly uneven (the ceiling score was 6.00). The same interpretation can probably be applied to the effect due to the interaction between Stimulus Imagery and Recall Trials which yielded $\underline{F}$ (1,104) = 11.98, $\underline{p} < .001$; the high-I stimulus pairs were recalled at a near ceiling level during the first recall trial. Nevertheless, these findings indicate that the effect of imagery is readily demonstrable during the initial stages of learning.

The interaction between Presentation Order and Response Imagery yielded $\underline{F}$ (1,104) = 5.05, $\underline{p} < .05$. This finding indicates that, in the response position, noun imagery is a more critical factor than adjective imagery. In contrast, the effect due to the interaction between Presentation Order and Stimulus Imagery was not significant ($p > .05$). Thus, imagery is a more important variable in the stimulus than in the response element of a word pair, whether the stimulus is a noun or adjective. This result shown in Figure 1 corroborates the findings of Paivio and his associates (Yuille, Paivio, & Lambert, 1969; Paivio, 1970).

Further support for the "stimulus peg" notion is provided by analyses of the effect due to the first order interaction between Stimulus Imagery and Response Imagery which yielded F (1,104) = 20.25, $p < .001$. The order of difficulty for learning under the various stimulus-response imagery conditions (from easiest to most difficult) was High-I Stimulus - High-I Response ($\overline{X} = 4.95$), High-I Stimulus - Low-I Response ($\overline{X} = 4.25$), Low-I Stimulus - High-I Response ($\overline{X} = 3.38$), and Low-I Stimulus - Low-I Response ($\overline{X} = 3.29$). These results indicate that increasingly greater gains were demonstrated as the pairs increased in concreteness. Thus, a High-I response element had a greater facilitative effect when the stimulus element was also a High-I noun or adjective. However, the significant triple interaction between the above factors and recall trials provides further qualification of this conclusion. This interaction, graphically displayed in Figure 2, is mostly attributable to the previously discussed ceiling effect that occurred in Trial 2. However, another differential effect between trials occurred in the ordering of the means where the

125

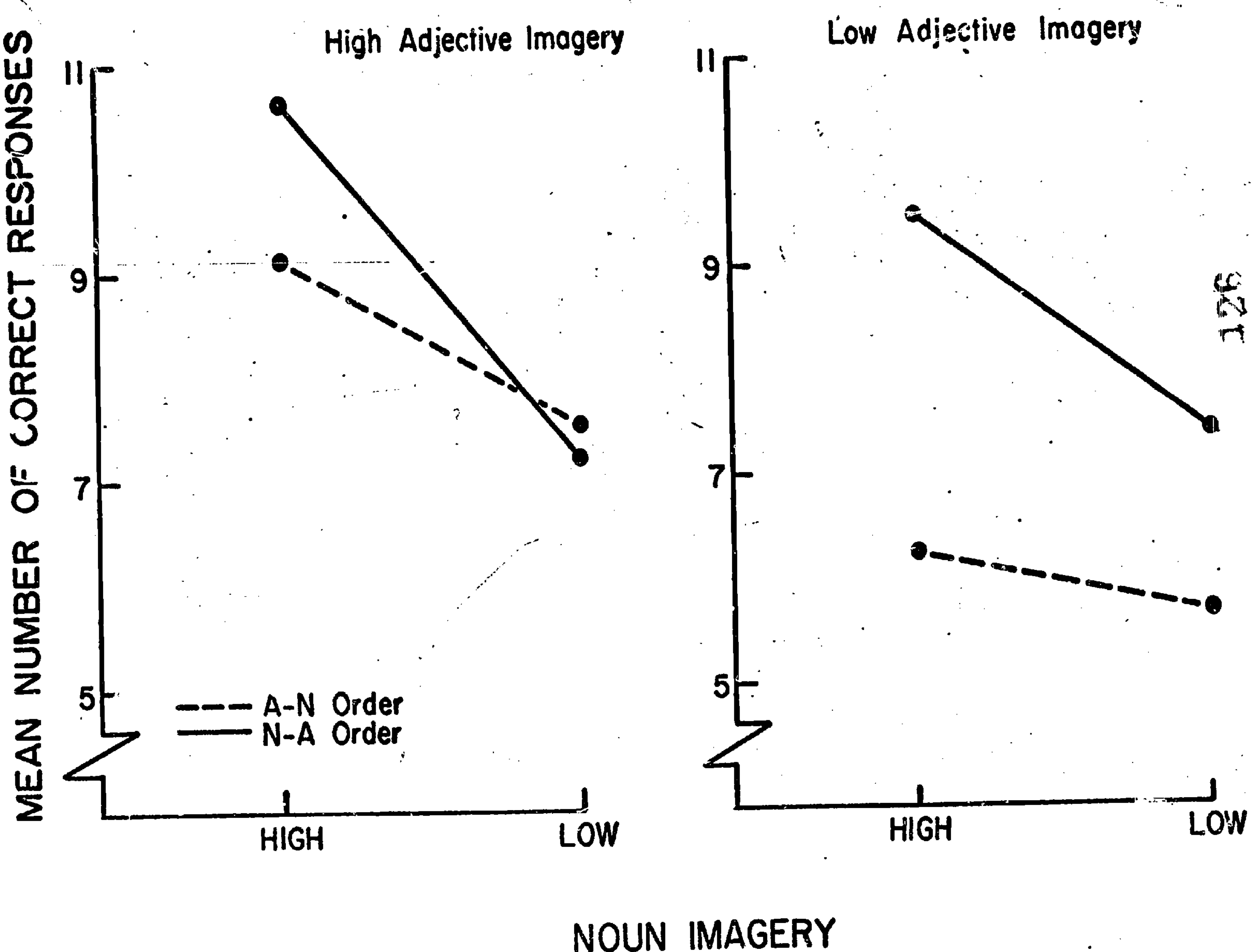

Figure 1. Mean number of correct responses for trials 1 and 2 combined, as a function of noun imagery, adjective imagery, and word order.

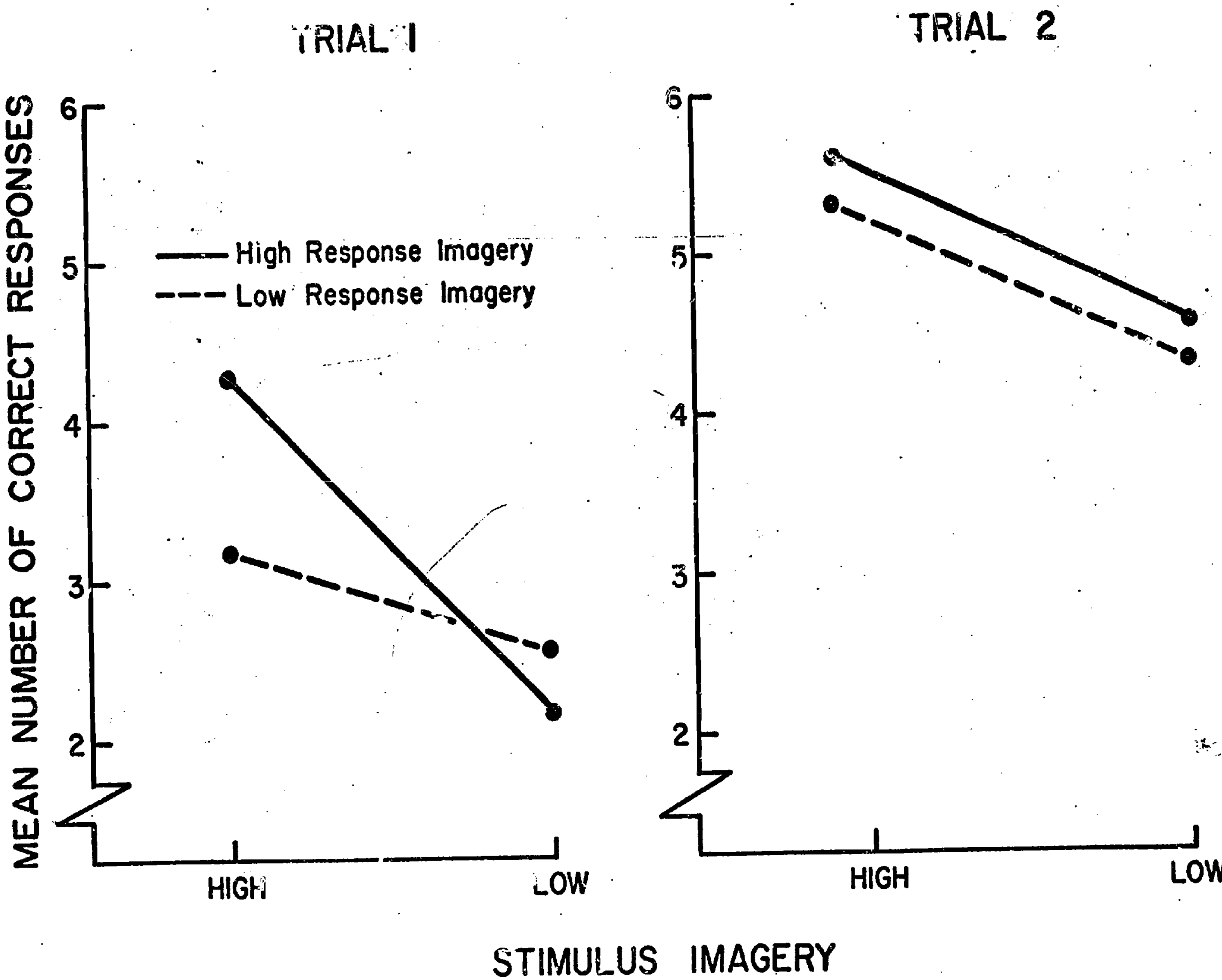

Figure 2. Mean number of correct responses for trials 1 and 2, as a function of stimulus imagery and response imagery.

hypothesized superiority of Low-I - High-I pairs over Low-I - Low-I pairs was reversed in trial 1, although they were not significantly different ($t < 1.00$). A subsequent analysis was performed on the data in the cells described above to test directly the hypothesis that imagery is more effective on the stimulus side than on the response side of paired-associates. In effect, all pairs with High-I stimulus elements ($\overline{X} = 3.73$ in Trial 1 and $\overline{X} = 5.48$) in Trial 2) were compared with those containing Low-I stimuli ($\overline{X} = 2.37$ in Trial 1 and $\overline{X} = 4.45$ in Trial 2), and similarly for the High-I ($\overline{X} = 3.22$ in Trial 1 and $\overline{X} = 5.12$ in Trial 2) and Low-I ($\overline{X} = 2.88$ in Trial 1 and $\overline{X} = 4.81$ in Trial 2) response elements. The overall comparison indicated that in both Recall Trial 1 ($t = 9.62$, $p < .001$) and Recall Trial 2 ($t = 6.57$, $p < .001$) stimulus imagery was a more critical determinant of recall facility than was response imagery.

Imagery aptitude was involved in a significant four-way interaction with Presentation Order, Stimulus Imagery, and Recall Trials [$F(1,104) = 4.75$, $p < .05$]. A summary of the means related to this interaction is presented in Table 2. Although the effects are not extreme, it is apparent that High Imagery ability was most effective in the recall of Low-I stimulus pairs. This result was more apparent in the A-N order during Trial 1 and in the N-A order during Trial 2. It is difficult to account for this order reversal between trials, but it is probably of questionable theoretical significance when the ceiling effect described above is taken into consideration. It should be noted that the performance of Low Imagers surpasses that of High

Table 2

Mean Number of Correct Responses on Trial 1 and Trial 2

As a Function of Imagery Aptitude, Presentation Order, and Stimulus Imagery

Stimulus Imagery and Imagery Aptitude	Presentation Order			
	Adjective-Noun		Noun-Adjective	
	Trial 1	Trial 2	Trial 1	Trial 2
Concrete Stimulus				
High-Imagers	3.02	5.37	4.56	5.81
Low-Imagers	3.30	5.00	4.07	5.72
Abstract Stimulus				
High-Imagers	2.02	4.37	2.60	4.98
Low-Imagers	1.57	3.98	2.46	4.48

Imagers when the High I stimulus was an adjective (A-N order). Thus, the data directionally support the hypothesis that the superiority of High Imagers over Low Imagers is greater when the stimulus is Low I. Additional support for this hypothesis was provided by the interaction between Imagery, Stimulus Imagery, Response Imagery and Recall Trials which yielded F (1,104) = 2.62, $p > .10 < .20$. Though this interaction was not significant, and indicates only a tendency, it suggests that the superiority of High Imagers over Low Imagers was most pronounced for Low-I - High-I pairs in Trial 1 (mean difference ($\bar{d}$) = +.58), and for Low-I - Low-I pairs in Trial 2 ($\bar{d}$ = +.51). On the other hand, the smallest differences between the two groups appeared in the learning of High-I - High-I pairs in Trial 1 ($\bar{d}$ = -.02) and also in Trial 2 ($\bar{d}$ = +.18). None of the other main effects or interactions were found to be significant.[1]

Discussion

The results of the present study provide clear support of Yuille, et. al's (1969) findings that imagery of nouns influences paired-associate learning more than does imagery of adjectives. With high imagery adjectives, nouns in the N-A order were more influential than were nouns in the A-N order. However, it is interesting that low imagery nouns coupled with high imagery adjectives were more effective in facilitating performance in the A-N than in the N-A order. When low imagery adjectives were employed, the N-A order was more effective than the A-N order whether high or low imagery nouns were used. These results imply that it is the rated imagery of the stimulus member of the pair that is most influential in its effect on rate of learning and retention. It is the concreteness of the stimulus rather than its form class that is the important variable. In contrast, the effect of meaningfulness of the response member has been found to be more influential in facilitating performance (Underwood & Schulz, 1960).

Although the results concerning individual differences in imagery ability were only suggestive, they are of theoretical importance. As Ernest and Paivio (1969) indicate, "Whether the imagery hypothesis may be extended to encompass individual differences in imagery ability is of considerable theoretical import. Successful predictions of performance based on such differences would provide further convergent evidence that a common intervening process is involved whether imagery is defined by stimulus attributes, mnemonic instructions, or individual differences" (p. 181). The present data indicate that high-imagers have an advantage over low-imagers when the stimulus is of low rated imagery. On the other hand, there is less difference in performance between the two groups when the stimulus is of high rated imagery. From these data, it appears that imagery ability affects performance for the same reasons that concrete stimuli do. When stimuli are concrete there is no further advantage to be gained by having high imagery ability; the stimuli are equally discriminable to both high and low imagers. However, when stimuli are of low-rated imagery, that is, when they are abstract, their ambiguity can be lessened by the imaginal ability of the high imager.

Though these data by themselves are of marginal significance, they gain importance when coupled with the results of earlier studies. Thus, Ernest and Paivio (1969) found that incidental recall was consistently better for high imagers than for low imagers. They (Ernest & Paivio, 1971) also found that, as measured by reaction times to elicit a verbal associate or to arouse an image, the high imager's performance was superior (i.e., latencies were shorter) to that of the low imager when the stimuli were abstract. The findings of the present study support this result. Of further support to this hypothesis is the study by Paivio & Foth (1970) whose <u>Ss</u> were <u>required</u> to either write sentences or draw pictures for a verbal mediation condition, or an imagery mediation condition, respectively. They found that verbal mediation facilitated the learning of abstract pairs, while imagery facilitated the learning of concrete pairs. However, in another condition, the Ss were merely provided mediation instructions (i.e., to generate either verbal or imaginal mediators) but were not required to employ sentences or drawings. Under this circumstance imagery was found to be better than verbal mediation for abstract pairs. This finding suggests that abstract pairs can be more easily learned with the aid of imagery processes. Thus, if imagery instructions are available but not forced upon S (Paivio & Foth, 1970) or if S has high imagery ability (Ernest & Paivio, 1971) there will be a positive effect upon abstract pair learning. Other evidence suggests that imagery ability may affect learning of abstract pairs differently from the recall of pictorial or verbal stimuli. Thus, Stewart (1965) and Kuhlman (1960) found that high-imagers recalled more items presented in pictorial than in verbal form, while the low-imagers recalled more verbal than pictorial items. Accordingly, it is hypothesized that acquisition and retrieval strategies generated in the free-recall task employed by Stewart and Kuhlman differ from those generated in a word-association or paired-associate learning task. Clustering and subjective organization based on pictures versus words in a free-recall task implies a preference for a given strategy (i.e., imaging) over another (e.g., employing verbal mediators). On the other hand, the effect of imagery in paired-associate learning implies the ability of the individual to employ a strategy that efficiently transforms the stimulus to a form necessary for effective hook-up with the response. Although imagery ability appears to be functionally related to learning and memory there is still insufficient evidence to indicate that this ability reflects the same process as that suggested by experimental manipulations of word imagery (Ernest and Paivio, 1969, p. 182). Nevertheless, the data from the present experiment strongly suggest that this may be the case.

Footnotes

[1] In a previous analysis Presentation List was included as a factor. It was not significant as a main effect but was involved in a significant interaction with Stimulus Imagery ($p < .05 > .01$) and in a five way interaction with Presentation Order, Individual Aptitude, Recall Trials, and Response Imagery ($p < .05 > .01$). Because the items in the list were selected at random there was no explanation for these differences. Inasmuch as the differences were not disordinal (in the interaction with Stimulus Imagery) or systematic in the five-way interaction, these interactions were disregarded in subsequent analyses.

References

Bennett, G. K., Seashore, H. G., & Wesman, A. G. Differential Aptitude Tests. New York: The Psychological Corporation, 1947.

Ernest, C. H., & Paivio, A. Imagery ability in paired-associate and incidental learning. Psychonomic Science, 1969, 15, 181-182.

Ernest, C. H., & Paivio, A. Imagery and sex differences in incidental recall (Research Bulletin No. 121) London Ontario: Department of Psychology, University of Western Ontario, 1970.

Ernest, C. H., & Paivio, A. Imagery and verbal associative latencies as a function of imagery ability. Canadian Journal of Psychology, 1971, 17, 83-90.

Kuhlman, C. K. Visual imagery in children. Radcliffe College, Unpublished doctoral thesis, 1960.

Kusyszyn, I., & Paivio, A. Transition probability, word order, and noun abstractness in the learning of adjective-noun paired-associates. Journal of Experimental Psychology, 1966, 71, 800-805.

Lambert, W. E., & Paivio, A. The influence of noun-adjective order on learning. Canadian Journal of Psychology, 1956, 10, 9-12.

Lockhart, R. S. Retrieval asymmetry in the recall of adjectives and nouns. Journal of Experimental Psychology, 1969, 79, 12-17.

Paivio, A. Learning of adjective-noun paired-associates as a function of adjective-noun word order and noun abstractness. Canadian Journal of Psychology, 1963, 17, 370-379.

Paivio, A. Mental imagery in associative learning and memory. Psychological Review, 1969, 76, 241-263.

Paivio, A. On the functional significance of imagery. Psychological Bulletin, 1970, 6, 385-392.

Paivio, A., & Foth, D. Imaginal and verbal mediators and noun concreteness in paired-associate learning: the elusive interaction. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 384-390.

Paivio, A., Yuille, J. C., & Madigan, S. A. Concreteness, imagery, and meaningfulness values for 925 nouns. Journal of Experimental Psychology, 1968, 76, (No. 1, Pt. 2).

Stewart, J. C. An experimental investigation of imagery. Unpublished doctoral dissertation, University of Toronto, 1965.

Thurstone, L. L. A factorial study of perception. Chicago: University of Chicago Press, 1944.

Thurstone, L. L., & Jeffrey, T. E. Flags: A Test of Spatial Thinking. Chicago: Industrial Relations Center, 1956.

Underwood, B. J., & Schulz, R. W. Meaningfulness and verbal learning. New York: Lippincott, 1960.

Yuille, J. C., Paivio, A., & Lambert, W. E. Noun and adjective imagery and order in paired-associate learning by French and English subjects. Canadian Journal of Psychology, 1969, 23, 459-466.

The Retrieval of Abstract and Concrete Materials By High and Low Imagers Employing Imaginal and Verbal Mediation With Abstract and Concrete Mnemonic Aids

Francis J. Di Vesta and Phyllis M. Sunshine

On the basis of a review of an extensive number of studies, Paivio (1969b) has suggested that a two stage model was implied for associative learning of noun pairs. Thus, the meaning of concrete nouns is acquired through both direct experience with the referent and association with other words. The consequence is that either verbal associations or nonverbal images might be evoked by these nouns to influence learning and recall. The meanings of abstract words, on the other hand, are acquired primarily, if not exclusively, through intraverbal experience. Accordingly, abstract words tend to elicit verbal rather than imaginal associations. Stated in a slightly different manner, the learner confronted with a task consisting of associating concrete nouns might employ both verbal and imaginal processes or strategies. However, either because he has a preference for using images or because images are more available than words, the learner tends to use the imagery strategy with concrete words. This conclusion is supported by both the subjects' subjective reports and by comparison of learning scores under the two strategies. Confronted with a task in which he is required to associate abstract terms, the learner employs the strategy emphasizing verbal association.

While a number of experimental procedures have been used to test these hypotheses, two methods in particular are of concern in the present study. One of these methods employs the "one is a bun" mnemonic device for facilitating recall. The jingle in this mnemonic aid can be composed of concrete pegwords, in which case it should facilitate the use of an imagery strategy, or it can be composed of abstract pegwords thereby facilitating the use of a verbal strategy. In the other research method of concern here instructions are employed to induce the subject to use either imagery or verbal associations when associating the word-pairs. On occasion both repetition-set and no-set conditions have been incorporated into the design as controls (e.g., Paivio & Yuille, 1969). Inasmuch as comparison with the latter two treatments indicate a clear superiority of verbal and imaginal processes for learning abstract and concrete pairs, respectively, the control conditions were not considered further for the present study.

Induction of these sets, through instructions only, does not always provide a strong effect. Accordingly, the procedure had been modified by Paivio & Foth (1970) by requiring the subject, in the imaginal set treatment, to draw a picture linking the pegword to its serial counterpart in the list to be learned; and by requiring the subject, in the verbal set treatment, to write a sentence using the two words. The importance of these methodological variations for the present investigation is that a means by which strategies can be manipulated is provided. Since both procedures influence the use of strategies in the same way (which, incidentally, is comparable to the effect of concreteness-abstractness described above) it was hypothesized that the use of the two methods in a crossed design, where the subject's task was to learn concrete and abstract lists, would result in further enhancement of the effect.

By employing these two treatments in conjunction with groups of high and low imagers it was hypothesized that the strategies used by subjects would become explicit in the ability of the two groups to learn and recall concrete and abstract lists. More specifically, it was hypothesized that since both verbal and imaginal processes could be used effectively with concrete materials, both groups of subjects would perform equally well on the task (see Ernest & Paivio, 1971). However, since the abstract list was assumed to be more easily associated to other words by verbal processes, the performance of the low-imagers would be especially hindered when forced to employ imaginal processes with the abstract mnemonic in learning an abstract list. Thus, the present study was an attempt to explore the possibility that "... the [effects of the] three classes of independent variables [in studies of imagery] - stimulus attributes, experimentally manipulated mediators, and individual differences [in imagery-ability]-... are mediated by common intervening processes" (Paivio, 1969b, p. 259).

Method

Design

The overall design of this study required that the subjects first learn (memorize) a jingle which was to serve as a conceptual peg for later learning tasks. Half of the subjects learned a jingle in which the critical words were concrete; the remaining subjects learned a jingle in which the critical words were abstract. These two treatments were orthogonally crossed with two mediational modes. Thus, when using the jingle as a mnemonic aid half of the subjects were required to employ it within an imagery-set, that is, they were to literally draw pictures, while the other half were to employ the jingle in a verbal set, that is, they were to write sentences to link the elements of the jingle with the new material to be learned. The jingles and sets were used to learn four different lists of words, two of which were comprised of concrete nouns and the other two were comprised of abstract nouns. Tests for recall of each list were administered immediately after learning the list and tests for recall of words in all lists were administered at the conclusion of the experiment. The subjects were selected on the basis of their imagery scores with half being high imagers and the other half low imagers. Where all variables were incorporated, the design implied a mixed analysis of variance with three between-subjects variables and one within-subjects variable.

Subjects

The subjects were 160 undergraduate students enrolled in the introductory educational psychology course at The Pennsylvania State University. They received credit toward their course grade for participating in the experiment.

Selection of Imagery Groups

Prior to the conduct of the experiment proper, 345 subjects were administered the Space Thinking (Flags) test (Thurstone & Jeffrey, 1959), the Space Relations test of the Differential Aptitude Test Battery (Bennett, Seashore & Wesman, 1963), and the Gottschaldt Figures Test as described by Thurstone (1944). A factor score for each subject was obtained following the procedure described by Glass and Maguire (1966) in which the raw test scores are weighted by their respective factor loadings. These loadings were obtained from a factor analysis conducted previously and described in an earlier report by Di Vesta, Ingersoll, & Sunshine (1971, in press). The 80 subjects with the highest factor scores were characterized as the high imagers and those 80 with the lowest scores were characterized as low imagers. The subjects within each of these groups were randomly assigned to each of the four experimental treatments involved in the mnemonic aid variable crossed with the mediational mode variable. The only restriction in the random assignment of subjects was that there would be an equal number of subjects (n = 20) in each cell of the design.

Materials

Serial learning lists. The 40 words used to construct the lists for the serial learning tasks were selected from the concreteness (C), imagery (I), and meaningfulness (*m*) norms reported by Paivio, Yuille, & Madigan (1968). Of the words selected, 20 were abstract and low on rated imagery and 20 were concrete and high on rated imagery. The two groups of words were equivalent in *m*. The means of the different attributes for the concrete (C) and abstract (A) lists, respectively, were: $\overline{X}$ = 6.54 and $\overline{X}$ = 2.90 for imagery-ratings; $\overline{X}$ = 6.91 and $\overline{X}$ = 2.11 for concreteness-ratings; and $\overline{X}$ = 6.13 and $\overline{X}$ = 5.50 for *m*, each of these latter two means is SD = ±0.3 from the mean *m* value of all words in Paivio, et al.'s norms. Two lists, of ten concrete nouns in each list were constructed by randomly selecting words from the initial list of 20 concrete nouns. A similar procedure was employed in constructing two lists of ten abstract nouns. The lists are presented in Table 1.

Table 1

Rated Imagery, Concreteness, and *m*-values for all Words In the Concrete and Abstract Serial Learning Lists

Word	Imagery Latency	Concreteness Rating	Meaningfulness (*m*)
List 1 - Concrete			
Piano	6.70	6.85	6.40
Candy	6.63	6.56	6.39
Accordian	6.50	7.00	5.89
Steamer	6.53	6.94	6.32
Dress	6.53	6.93	5.68
Elbow	6.30	6.94	5.16
Mule	6.60	6.96	6.12
Cigar	6.80	6.96	6.22
Frog	6.73	6.96	6.56
Macaroni	6.47	7.00	5.48
Mean	6.58	6.91	6.02
List 2 - Concrete			
Word			
Library	6.73	6.87	6.40
Skull	6.47	6.96	6.64
Tweezers	6.57	6.93	5.80
Engine	6.33	6.76	6.08
Corpse	6.50	6-89	6.52
Building	6.40	6.94	5.48
Headlight	6.43	6.90	6.32
Pipe	6.43	6.90	6.20
Leopard	6.77	7.00	6.83
Nail	6.50	6.96	6.08
Mean	6.51	6.91	6.24
Average of Concrete Lists	$\overline{X}_I = 6.54$	$\overline{X}_C = 6.91$	$\overline{X}_m = 6.13$

Table 1a

Rated Imagery, Concreteness, and m-values for all Words

In the Concrete and Abstract Serial Learning Lists

Word	Imagery Latency	Concreteness Rating	Meaningfulness (m)
List 3 - Abstract			
Adversity	2.80	2.03	5.06
Belief	2.73	1.55	5.24
Ego	2.90	1.93	5.72
Irony	2.83	2.10	5.25
Rating	2.60	2.66	5.12
Hypothesis	2.40	2.25	5.36
Emancipation	3.20	2.49	5.20
Deceit	3.30	1.66	4.92
Exclusion	2.80	2.41	5.32
Ability	2.67	2.03	5.60
Mean	2.82	2.11	5.28
List 4 - Abstract			
Word			
Crisis	3.43	2.81	5.44
Mercy	3.40	1.59	5.20
Satire	3.37	2.33	5.67
Magnitude	2.50	3.03	5.68
Knowledge	2.97	1.56	6.36
Perception	3.17	2.33	5.80
Democracy	2.47	1.79	5.72
Intellect	2.93	1.83	5.56
Welfare	3.17	2.35	6.16
Chance	2.50	1.51	5.61
Mean	2.99	2.11	5.72
Average of Abstract Lists	$\bar{X}_I = 2.90$	$\bar{X}_C = 2.11$	$\bar{X}_m = 5.50$

Jingle words. The concrete and abstract jingle words were selected from a pool of 25 words rhyming with the numbers one to ten. The initial pool of words was obtained by searching two dictionaries (Stillman, 1965; Wood, 1936) of rhyming words. The nouns were rated by 15 subjects for the ease with which they evoked sensory images on a seven-point scale (Paivio, et al., 1968). Concreteness was also rated on a seven-point scale bounded by the terms abstract and concrete. Concreteness was defined in terms of the directness of sensory reference as used by Spreen & Schulz (1966). The production of data regarding meaningfulness (m) was accomplished by requiring 15 subjects to associate as many words as possible, within 30-secs., to each noun. The procedure described by Noble (1952) was used to obtain the m values for these words.

The mean C and I ratings of the nouns selected for the jingle with concrete pegwords were $\bar{X} = 6.06$ and $\bar{X} = 6.72$, respectively; while the mean C and I ratings of the nouns selected for the jingle with the abstract pegwords were $\bar{X} = 1.86$ and $\bar{X} = 2.52$, respectively. The mean m of the concrete nouns was 6.05 and of the abstract nouns it was 5.95. Thus, the jingle words differed on the basis of imagery and concreteness but were essentailly the same on the basis of meaningfulness.

The words selected for the jingle comprised of concrete nouns were: one-bun; two-shoe; three-tree; four-door; five-hive; six-sticks; seven-leaven; eight-gate; nine-wine; and ten-hen. The words selected for the jingle comprised of abstract nouns were: one-fun; two-review; three-spree; four-chore; five-tithe; six-rhetoric; seven-heaven; eight-fate; nine-divine; ten-amen.

Procedure

The subjects were administered the experimental tasks in groups of two to four. At the outset, they were instructed in the reporting of verbal and imaginal mediation. Each person was given a stopwatch and instructed on its use in order to time himself at each task. After practice in using the stopwatch for two to three minutes, they memorized either the concrete or abstract jingle until the criterion of two successively correct recitations without error was reached. The experimenter then read aloud one of two sets of instructions depending upon the treatment being administered, i.e., the subjects were instructed to employ an imagery-set or a verbal set in performing the tasks. In the use of the imagery-set the subjects were instructed to connect, with a mental picture or image, each noun in the list to be learned with the jingle noun in the corresponding serial position and to record his connection by drawing a picture, however crude it might be. The verbal-set required subjects to relate, by forming a sentence or phrase, each noun in the list to the jingle noun in the corresponding serial position and to record his connection in <u>verbal</u> form. Subjects were reminded, before each list to be memorized had been read, to reset their stopwatches.

Following the preliminary instructions, the subjects were given two concrete noun-pairs and one abstract noun-pair in order to practice the use of the jingle and mediational set. The experimenter then read aloud the first noun in the list, and instructed the subjects to "start" at which time the watches were started. The subject stopped the watch as soon as he formed the complete mental image or verbalization. After the connection was recorded the subject recorded the time to the nearest second needed to form the link-up. The subjects were tested for serial recall after each of the four lists was presented.

The presentation of the four lists were counterbalanced among subjects by the use of a simple Latin-square to minimize the possible effects on recall of the order in which the lists were presented. After the recall test for the last list the experimenter asked the subjects to recall all of the words from the four lists that were in the first position, all of the words from the four lists that were in the second position and so on to the words in the tenth position, to determine how many of the 40 words the subject could retrieve.

Results

Several measures of performance were obtained including number of concrete- and abstract-word omissions after each list was memorized, intrusions from one list to another in recall sessions immediately after each list, omissions in the final recall task, intrusions from one list to another in the final recall task, and latencies in arriving at an association during the memorization task itself. Overall intercorrelations between the imagery-ability measure and each of these dependent variables indicated relatively high interrelationships among the measures. Accordingly, it was decided that the most efficient procedure was to perform analyses of the latencies in arriving at an association between the mnemonic aid (jingle) and the words in the list to be memorized; total errors, separately for concrete and abstract lists, made during the recall tasks <u>after</u> each list; and total errors of both types made on the final recall task. Mixed analyses of variances were made of each measure. In each analysis the between-subjects variables were level of imagery-ability (high or low imagers), kind of mediator (pictorial or verbal), kind of mnemonic aid (concrete or abstract pegwords). The within-subjects variable was kind of list (concrete words or abstract words) to be learned. The analyses based on these measures are summarized in Table 2. Since a triple-interaction involving imagery-ability, mediational set, and kind of list was obtained in the analysis of errors made on the final recall task, separate factorial analyses of variance, based only on the between-subjects variables, were routinely made for error scores on the concrete lista and for error-scores on the abstract lists of words. The results of these subanalyses are reported below only where it seemed necessary to do so in order to clarify the locus of a given effect.

Latency Measures

The initial analysis involved the measure of time, in seconds, to arrive at an association between the pegword in the mnemonic aid and the corresponding word in the list to be memorized. This analysis yielded F (1,152) = 18.01, $p < .01$, for the main effect due to the kind of mediational set; F (1,152) = 10.11, $p < .01$ for the main effect due to the kind of mnemonic aid; and F (1,152) = 14.99, $p < .01$ for the main effect due to kind of list memorized.

The time taken to arrive at an association by subjects who were to use a picture (imaginal set) in linking the pegword with a list word was longer on the average ($\overline{X}$ = 138.14 secs.) than that taken by subjects who linked the two sets of words via a sentence or verbal set ($\overline{X}$ = 99.69 secs.). It took less time to link each of the concrete words ($\overline{X}$ = 104.52 secs.) with its corresponding pegword than it did to link each of the abstract words ($\overline{X}$ = 133.32 sec.).

Table 2

Summary of Analyses of Variance of Recall Errors

		Latency[a]		Errors Immediate Recall		Errors Final Recall	
Source	df	MS	F	MS	F	MS	F
Between-subjects							
Imagery (A)	1	18942	2.89	147.15	8.97*	277.51	11.18*
Mediators (B)	1	118272	18.01*	.70	.04	4.05	0.16
Mnemonic (C)	1	66355	10.11*	302.25	18.42*	515.11	20.76*
A x B	1	7527	1.15	169.65	10.34*	357.01	14.39*
A x C	1	22916	3.49**	11.63	0.71	2.45	0.10
B x C	1	588	0.09	104.65	6.38**	143.11	5.77**
A x B x C	1	5461	0.83	48.83	2.98	115.20	4.64**
Error (b)	152	6566		16.41		24.82	
Within-subjects							
Lists (J)	1	25920	14.99*	580.50	64.76*	1748.45	256.57*
A x J	1	2464	1.43	0.08	0.01	2.11	0.31
B x J	1	2565	1.48	24.75	2.76	20.00	2.94
C x J	1	610	0.35	0.53	.06	78.01	11.45*
A x B x J	1	485	0.28	2.28	.25	37.81	5.55**
A x C x J	1	262	0.15	0.38	.04	4.50	0.07
B x C x J	1	2952	1.71	1.38	.15	3.12	0.05
A x B x C x J	1	768	0.44	1.40	1.57	0.00	0.00
Error (w)	152	1729		8.96		6.81	

[a] Decimal places have been dropped for mean squares of analysis of latency scores.

* $p < .01$

** $p < .06$

The effect due to the interaction between imagery ability and kind of mnemonic aid used yielded F (1,152) = 3.49, p < .06. The means related to the interaction between imagery-ability and kind of mnemonic aid employed indicated the difference in time taken by high imagers ($\overline{X}$ = 105.29 secs.) and low imagers ($\overline{X}$ = 103.75 secs.) when using the concrete mnemonic was not significant. However, when the high imagers used the abstract mnemonic they took an average of 117.16 secs. for linking each words to its conceptual peg while the low imagers required an average of 149.47 secs. This finding is consistent with predictions form Paivio's two-stage association model.

The main effect due to imagery and the remaining interactions were found not to be significant (p > .10).

Errors on the Immediate Recall Task

The total number of errors made on recall after each list were analyzed via a mixed analysis of variance. The score for the number of errors was based on the summed omissions and intrusions. Thus, the score represents the exact opposite of the number of correct responses. This analysis indicated that the difference between imagery-ability groups was significant, F (1,152) = 8.97, p < .01; that there was a significant main effect due to mnemonic aids, F (1,152) = 18.42, p < .01; and a significant main effect due to kind of list memorized, F (1,152) = 64.76, p < .01. In addition, the analysis revealed significant interactions between imagery-ability and kind of mediator, F (1,152) = 10.34, p < .01; between mediators and mnemonic aids, F (1,152) = 6.38, p < .05.

The differences in mean number of errors among groups indicated in the aforegoing analyses, are summarized in Table 3. In brief, these data indicate that high imagers made significantly fewer errors ($\overline{X}$ = 3.36) than low imagers ($\overline{X}$ = 6.17) when imaginal sets were used. However, the difference in means for the two groups ($\overline{X}$ = 4.73 and $\overline{X}$ = 4.62 for high and low imagers, respectively) was not significant when verbal mediators were used. Additionally supportive of the notion that imagery facilitates learning and recall is the finding that fewer errors were made with concrete mnemonic aids ($\overline{X}$ = 3.75) than with abstract mnemonic aids ($\overline{X}$ = 5.69) and that this difference is considerably greater when imaginal mediators (difference = 3.09) were used than when verbal mediators (difference = 0.79) were used. There is a hint in this analysis of the interaction between imagery and treatments which appears strongly in the analysis of the final recall data presented immediately below. Thus, in the present analysis high imagers made fewer errors ($\overline{X}$ = 2.40) than low imagers ($\overline{X}$ = 4.05) when imaginal mediators were used with concrete mnemonic devices, t = 1.28, p < .10. Low imagers were especially hindered ($\overline{X}$ = 8.30) when using imaginal mediators and abstract mnemonic devices conjunctively compared to high imagers ($\overline{X}$ = 4.33), t = 3.09, p < .01. When using the verbal set with concrete mnemonic aids high imagers tend to make slightly fewer errors ($\overline{X}$ = 4.13) than low imagers ($\overline{X}$ = 4.42); and when a verbal set is used with abstract mnemonic-aid high imagers make more errors ($\overline{X}$ = 5.33) than do low imagers ($\overline{X}$ = 4.83). However, the interaction represented by these means is not significant (p > .10).

Table 3

Summary of Mean Number of Errors Made on the Recall Task, Immediately After Each List Presentation, by High and Low Imagers Employing Concrete and Abstract Mnemonics Under Conditions of Imaginal and Verbal Mediation

Imagery-Ability	Kind of Mnemonic: Concrete Mnemonic	Abstract Mnemonic	Overall
Imaginal-set			
High Imagers	2.40	4.33	3.36
Low Imagers	4.05	8.30	6.17
Total	3.22	6.31	4.76
Verbal-set			
High Imagers	4.13	5.33	4.73
Low Imagers	4.42	4.83	4.62
Total	4.28	5.07	4.67
Overall	3.75	5.69	4.72

151

Errors on the Final Recall Task

The analysis of errors in the final recall task was based on the score involving all incorrect responses and omissions made in recalling words in given positions from all lists after the four lists had been memorized. In this analysis the effect due to imagery-ability yielded F (1,152) = 11.18, $p < .01$; that due to kind of mnemonic aid yielded F (1,152) = 20.76, $p < .01$; and the effect due to lists yielded F (1,152) = 256.67, $p < .01$. Of the first order interactions three were found to be significant. These were F (1,152) = 14.39, $p < .01$, for the interaction between imagery-ability and kind of mediational set; F (1,152) = 5.77, $p < .05$ for the interaction between mediational set and mnemonic aids; and F (1,152) = 11.45, $p < .01$ for the interaction between kind of mnemonic aids and kind of lists. These main effects and interactions must be further qualified by the significant second-order interaction between imagery-ability, mediational set and mnemonic aids which yielded F (1,152) = 4.64, $p < .05$, and between imagery-ability, mediational set and kind of list which yielded F (1,152) = 5.55, $p < .05$.

The mean number of errors for these effects are summarized in Tables 4 and 5. The reader will note that high imagers made fewer errors ($\overline{X}$ = 5.60) than low imagers ($\overline{X}$ = 9.60) when using imaginal mediators. On the other hand, there is no difference between the two groups when verbal mediators are used, high imagers made an average of 7.96 errors while low imagers averaged 7.71 errors. The performance of high imagers was particularly facilitated when they used imaginal mediators with concrete mnemonic aids ($\overline{X}$ = 4.37) and concrete words ($\overline{X}$ = 3.80). The low imagers were hindered to a considerable degree when they employed imaginal mediators with the abstract mnemonic aid

Table 4

Summary of Mean Number of Errors Made During Final Recall
By High and Low Imagers Employing Imaginal and Verbal Mediators
In Learning Concrete and Abstract Lists

Imagery Ability	Kind of List: Concrete	Abstract	Row Means
<u>Imaginal-mediator</u>			
High Imagers	3.80	7.45	5.60
Low Imagers	7.25	11.95	9.60
Sub-means	5.53	9.70	7.61
<u>Verbal-mediator</u>			
High Imagers	4.95	10.98	7.96
Low Imagers	5.58	9.88	7.71
Sub-means	5.25	10.43	7.83
Column means	5.39	10.06	7.72

Table 5

Summary of Mean Number of Errors Made During Final Recall

By High and Low Imagers Employing Concrete and Abstract Mnemonics

With Imaginal and Verbal Mediators

Imagery Ability	Kind of Mnemonic: Concrete Mnemonic	Abstract Mnemonic	Row Means
Imaginal Mediator			
High Imagers	4.37	6.87	5.63
Low Imagers	6.97	12.22	9.60
Sub-means	5.67	9.55	7.61
Verbal Mediator			
High Imagers	6.85	9.08	7.96
Low Imagers	7.63	7.80	7.71
Sub-means	7.24	8.43	7.83
Column means	6.46	8.99	7.72

($\overline{X} = 12.22$) and with the abstract word-list ($\overline{X} = 11.95$). Multiple comparisons made of the data presented in Tables 4 and 5 via the t-test indicate only the differences between high and low imagers employing the imaginal mediator with abstract lists ($t = 3.31$) or with abstract mnemonics ($t = 3.38$) were significant ($p < .01$). The interaction between imagery ability and kind of list in the verbal mediation condition yielded $F = 1.87$, $p < .065$.

In general, high imagers always made fewer errors than did low imagers when imaginal mediators were used. High imagers retained this advantage, although to a considerably lesser degree, when verbal mediators were used with the concrete mnemonic and concrete word list. However, the low imagers made fewer errors than high imagers when learning the abstract word lists or when using the abstract mnemonic when verbal mediators were used. Thus, while the analysis of results on delayed recall are in general agreement with those obtained for immediate recall, the differences obtained are larger. Accordingly, it appears that one condition for identifying aptitude by treatment interactions is the examination of its influence of imagery, under conditions comparable to those in the present experiment, over longer delay periods, perhaps a delay of a week or two.

Discussion

The results of this study clearly imply that rated imagery (i.e., concreteness) as a stimulus attribute, imaginal strategy as a mediational process, and imagery-ability as an individual difference variable are similarly related to performance and recall. Thus, pairs of concrete nouns are acquired more easily than pairs of abstract nouns, an imaginal mediational strategy yields fewer errors than a verbal mediational strategy, and learners with high imagery-ability produce fewer errors than those with low imagery-ability.

The critical comparisons for this study, however, involved the interactions among these variables, particularly the performance of high and low imagers when learning word-pairs comprised of abstract stimuli or when using abstract mnemonics under sets to employ imaginal rather than verbal processes. With respect to these relationships it was found that for the immediate recall task, the high imagers performed significantly better than low imagers when the abstract mnemonic device was combined with the imaginal set. In no other condition was the high imagery group superior to the low imagery group on this task. The same result was obtained in the final (delayed) recall task. The concreteness of words within lists as a variable interacting with imagery ability was not supported when imaginal mediators were used; i.e., the low imagers made significantly more errors than high imagers on both kinds of lists when imaginal mediators were used. However, the interaction between imagery ability and kind of list on the final recall task indicated a tendency for high imagers to perform better on the concrete task and low imagers to perform better on the abstract task when verbal mediators were used. Furthermore, in all comparisons with either abstract word lists or mnemonic devices with abstract pegwords, low imagers performed significantly ($p < .01$) better with verbal sets than with imaginal sets.

It would appear that the two-stage model requires, and has received in this study, both kinds of support when comparing the performance of high and low imagers. Thus, high imagers when learning abstract stimuli impose an imagery strategy on the material being learned. When required (i.e., forced) to use this strategy they can perform effectively. However, because they typically employ an imaginal strategy in processing information, if required to use another strategy, such as a verbal one, they perform much more poorly than the low imagers who, presumably, employ other strategies to better advantage than the imagery one. Conversely, the low imager who does not employ imagery to good advantage has difficulty in using the imaginal set, in fact his performance is seriously hampered when compared to his performance under the verbal set. Thus, the hypotheses regarding the two-stage model and the role of imagery ability in associative learning are provided considerable support in the present study.

A major difficulty in conducting studies on trait by treatment interactions appears to be in determining how the trait is to be measured. Thus, in the present study, high imagers perform about as expected. However, low imagers are not really to be considered verbalizers. It is not clear what their dominant strategy is except that they are people deficient in some strategy (i.e., in imagery-ability) without knowing their strengths. Yet some means of identifying a group with a strategy that is "opposite" to imagery (verbalizers?) in the same way that imaginal processes are "opposed" to verbal processes is clearly required. Upon analysis, identification of a strategy such as that of verbalizing may be found to be a difficult task. Are verbalizers learners who are verbally fluent? with excellent vocabularies? with flawless language habits? with good reading comprehension? a composite of all of these? A general verbal trait or ability can be easily identified; a trait closely related to the acquisition of associates may be a much more difficult task.

Nevertheless, the overall results of this study clearly indicate that whatever has been measured by the battery of "imagery" tests is definitely related to the processing of information by the learner. Thus, the major distinction between high and low imagers is at least in their ability, or preference, to use imagery over some other learning-recall strategy. When conditions favor the use of this strategy the high imager is at an advantage. This general conclusion is supportive of the two-stage model of associative learning and is comparable to a result reported by Yuille & Paivio (1967) who found that mediation latency was unrelated to stimulus concreteness and mediation set when concrete stimuli were employed but that imagery set was significantly inferior to verbal set when abstract stimuli were used. In the present study the parallel comparison is of the performance of low imagers employing imaginal and verbal sets with abstract mnemonics or lists. In each of these comparisons, the performance of the low imager is hindered when he is forced to employ imaginal mediators, but is unimpaired when he is forced to use verbal mediators.

In summary, the reasoning and results of the present study indicate that the relationships between traits and treatments may be in the ability of the individual to deal with the task in general, his receptivity to, or preference for, certain kinds of stimuli over others, or the strategies by which he attacks a task or processes the information. Stewart (1965) and Hollenberg (1970) both assumed that imagery-ability affected the receptivity to stimuli and accordingly their investigations compared the acquisition and recall of learners when presented pictorial and verbal stimuli. Without minimizing the importance of presentation mode as a variable in learning, the present study suggests that the manner in which the learner processes the material, in terms of the task requirement (also see Ernest & Paivio, 1969), is as important as the effects of manner of presentation. For an understanding of the dynamics of the learning process including the role of individual differences on performance, it may be more important.

References

Bennett, G. K., Seashore, H. G., & Wesman, A. G. Differential aptitude tests. (Grades 8-13 and adults.) New York: Psychological Corporation, 1963.

Di Vesta, F. J., Ingersoll, G., & Sunshine, P. A factor analysis of imagery tests. Journal of Verbal Learning and Verbal Behavior, 1971, in press.

Ernest, C. H., & Paivio, A. Imagery ability in paired-associate and incidental learning. Psychonomic Science, 1969, 15, 181-182.

Ernest, C. H., & Paivio, A. Imagery and verbal associative latencies as a function of imagery ability. Canadian Journal of Psychology/Review of Canadian Psychology, 1971, 25, 83-90.

Glass, G. V., & Maguire, T. O. Abuses of factor scores. American Educational Research Journal, 1966, 3, 297-304.

Hollenberg, C. K. Functions of visual imagery in the learning and concept formation of children. Child Development, 1970, 41, 1003-1015.

Noble, C. E. An analysis of meaning. Psychological Review, 1952, 59, 421-430.

Paivio, A. Learning of adjective-noun paired associates as a function of adjective-noun word order and noun abstractness. Canadian Journal of Psychology, 1963, 17, 370-377. (a)

Paivio, A. Mental imagery in associative learning and memory. Psychological Review, 1969, 76, 241-263. (b)

Paivio, A., & Foth, D. Imaginal and verbal mediators and noun concreteness in paired-associate learning: the elusive interaction. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 384-390.

Paivio, A., & Yuille, J. C. Changes in associative strategies and paired-associate learning over trials as a function of word imagery and type of learning set. Journal of Experimental Psychology, 1969, 79, 458-463.

Paivio, A., Yuille, J. C., & Madigan, S. A. Concreteness, imagery, and meaningfulness values for 925 nouns. Journal of Experimental Psychology Monograph Supplement, 1968, 76, No. 1, Part 2.

Spreen, O., & Schulz, R. W. Parameters of abstraction, meaningfulness, and pronunciability for 329 nouns. Journal of Verbal Learning and Verbal Behavior, 1966, 5, 459-468.

Stewart, J. C. An experimental investigation of imagery. Unpublished doctoral dissertation, University of Toronto, 1965.

Stillman, F. The Poet's Manual and Rhyming Dictionary. Thomas Y. Crowell Company, New York, 1965.

Thurstone, L. L. A factorial study of perception. Chicago: University of Chicago Press, 1944.

Thurstone, L. L., & Jeffrey, T. G. Space thinking (Flags). Chicago: Education-Industry Service, 1959.

Wood, C. The Complete Rhyming Dictionary and Poets Craft Book. Halcyon House: New York, 1936.

Yuille, J. C., & Paivio, A. Latency of imaginal and verbal mediators as a function of stimulus and response concreteness-imagery. Journal of Experimental Psychology, 1967, 75, 540-544.

161

Summary

The Effect of Context Modality on Acquisition and Transfer By Imagers and Non-Imagers*

Study Director: Steven M. Ross

Advisor: Francis J. Di Vesta

Technical Problem

This study investigated the conditions under which individuals who differ in imagery abilities acquire and transfer concepts that are incidentally expressed by pictorial or verbal contexts. It has been shown in an earlier investigation (Di Vesta & Ross, 1970) that the relatedness or meaningfulness of a verbal context has demonstrable effects upon the learning and transfer of paired-associates. Specifically, it was determined that a related context, i.e., one which is similar in meaning or categorically relatable to the stimulus side of the pair, elicits conceptualizing tendencies in the learner that interfere with specific item learning but facilitate conceptual transfer.

The present study extended the earlier one by manipulating the modality of the context, and also by incorporating imagery ability as an individual difference variable. It was hypothesized that since imagers would be more receptive to pictorial than to verbal contextual cues,

---

* This is an abstract of a master's thesis in preparation at the time the present report was being prepared.

they would acquire the pictorially expressed concept more readily than non-imagers. Non-imagers, on the other hand, were expected to be less influenced by context modality and therefore able to learn concepts expressed either pictorially or verbally with equal facility.

General Methodology

The experiment consisted of two phases: learning and transfer. During the learning phase, subjects learned a list of paired-associates to a criterion of one errorless trial. In one condition, two context words that were meaningfully related to each other and to the specific stimulus elements, were placed between the stimulus and response positions of all word-pairs. In a second condition the context was identical in meaning and similarly positioned as in the first condition but was presented pictorially. In a third condition no context of any kind was present. All stimuli were mounted on slides and presented by means of a carousel projector.

During the transfer phase, subjects learned a new list of paired-associates. Each word-pair in the list was comprised of the original response element from the learning phase list and a new stimulus. Context words were eliminated in the transfer phase. In one set of conditions the new stimulus word was categorically related to the original stimulus and to the context presented during the learning phase. In the other set of conditions the new stimulus was related only to the original stimulus, but not to the context. Thus, depending upon condition, subjects were required to transfer to a concept that was either identical or alternate to that of the learning context.

Technical Results

The analysis for the learning phase of the experiment indicated that the aptitude-by-treatment interaction involving imagery ability and context modality was significant. The direction of the data revealed that there were no differences in recall between imagers and non-imagers in the verbal context condition but that imagers were highly superior in both the picture context and no-context conditions. There were no significant differences overall between the context variations.

The analysis of data for the transfer phase indicated that performance on the same-concept transfer task was significantly better than performance on the alternate-concept transfer task. More important, however, was the finding that this difference was not pronounced when the pictorial context, as opposed to verbal context, was employed in the learning phase. Imagery ability did not significantly interact with these factors though the positive effects of the picture context tended to be slightly greater for imagers than they were for non-imagers.

Educational Implications

The present investigation involved the assumption that some individuals, more than others, habitually employ concrete images in the encoding of informational inputs. Accordingly, it appears likely that these individuals would demonstrate a greater capacity for learning from concrete kinds of materials than they would from materials that were relatively symbolic or abstract.

Though the present findings cannot be interpreted as unconditionally supportive of the above notion, they clearly imply that modality of presentation and relatedness of incidental cues provide constraints on the storage and transfer of incoming information. It can also be inferred from this study that, at least during acquisition, high imagers are more receptive to and more able to effectively process information that is embedded within a pictorial context than are low imagers. This suggests that students who demonstrate a preference for coding via imagery might best benefit from a context of concrete examples and graphic displays. If this interpretation proves to be viable then it implies a definite responsibility for the teacher to adapt the mode of the instruction to the individual preference and ability of the learner provided economy of acquisition is an important objective.

Implications for Further Research

In the present study, the effects of imagery as a stimulus attribute and as an individual difference variable were examined as determinants of learners' abilities to acquire and transfer information. The differential performances of imagers and non-imagers under the influence of pictorial and verbal contexts implies the need to investigate further the adaptation of instructional techniques to this kind of individual difference. Of considerable importance would be the assessment of the generalizability of the above findings and the degree to which they are translatable to applied classroom practices. Further research might also suggest methods of instruction that provide remedial assistance for individuals who demonstrate an excessive reliance upon visual strategies of processing information.

References

Di Vesta, F. J., & Ross, S. M. Contextual cues and cognitive structures in the storage and retrieval of information. Semi-annual report submitted to the Advance Research Projects Agency, 1970. Pp. 68-97.

166